
Nancy-Université
Université
Henri Poincaré

Université Henri Poincaré – Nancy I
U.F.R. Sciences et techniques de la matière et des procédés
École doctorale lorraine de chimie et physique moléculaires



Thèse

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré

en Chimie informatique et théorique

par Vincent LEROUX

Modélisation d'inhibiteurs du domaine SH2 de la protéine Grb2 par dynamique moléculaire, docking et criblage virtuel

sous la direction de Bernard MAIGRET

Équipe de dynamique des assemblages membranaires, UMR UHP / CNRS 7565

Soutenue publiquement le 15 décembre 2006

Rapporteurs	Dr. Nohad GRESH	Université René Descartes – Paris V
	Dr. Didier ROGNAN	Université Louis Pasteur – Strasbourg
Examinateurs	Prof. Daniel CANET	Université Henri Poincaré – Nancy I
	Prof. Christiane GARBAY	Université René Descartes – Paris V
	Dr. Bernard MAIGRET	Université Henri Poincaré – Nancy I
Invités	Dr. Peter BLADON	Interprobe Chemical Services, Glasgow, Écosse
	Dr. Gilles MOREAU	30 avenue Jean Jaurès, 94220 Charanton

Mes remerciements les plus chaleureux vont à mon directeur de thèse, Bernard, pour son encadrement, sa patience face à mon fonctionnement irrégulier, mais surtout pour avoir fait en sorte que je termine cette thèse en ayant une vision de la recherche scientifique qui a dépassé le simple stade de l'intérêt et de la curiosité. Un grand merci également à Nohad, Christiane, Peter et Gilles pour leur soutien et leur gentillesse. Je tiens également à saluer l'ensemble des membres du laboratoire eDAM, ainsi qu'à tous ceux avec qui j'ai travaillé, spécialement mes compagnons d'armes Alexandre, Jean-Paul, Matthieu, Jérôme, Adil, Amel, Yesmine, Muhannad et Werner. Bon courage à ceux qui n'en ont pas encore fini, et pareillement à ceux qui sont partis vers d'autres aventures. Enfin, je garde beaucoup de reconnaissance pour ma famille et mes amis, qui ont su rester présents même lorsque j'étais un peu déconnecté par mon travail et la distance. Merci en particulier à mes parents, Sophie, Pascal, Estelle et Catherine. Je ne regrette pas une seule seconde d'avoir fait cette thèse, cette expérience m'a apporté beaucoup, et chacun d'entre vous m'y a aidé.

Note importante

Ce document constitue une version révisée du manuscrit de thèse.

Il répercute les suggestions formulées par le jury lors de la soutenance.

La section publications a été modifiée, car au moment de l'impression du manuscrit de thèse les articles n'étaient pas publiés – deux d'entre eux étaient toutefois sous presse. On trouvera ici les publications dans leur version finale. La liste des changements intervenus pour l'un d'entre eux pourra être consultée en exergue de la section publications.

Ce document (ainsi que sa révision initiale) est disponible sous forme électronique sur demande : leroux.vincent@free.fr

---

Table des matières

Présentation

L’application de simulations numériques reposant sur les bases théoriques de la chimie est une approche relativement récente dans le monde de la recherche pharmaceutique. Plusieurs facteurs la rendent particulièrement attrayante.

Tout d’abord, son champ d’application croît parallèlement aux progrès réguliers de la puissance informatique accessible – une croissance bien plus spectaculaire que celle des avancées technologiques au niveau des appareillages de la recherche biomédicale classique. Récemment, de nouveaux concepts de calcul distribué, popularisés par le projet SETI@Home, ont encore accru l’intérêt de la communauté scientifique envers les approches *in silico* en général.

Ensuite, l’utilisation de modèles numériques peut permettre, lorsque ceux-ci se basent sur la structure des assemblages moléculaire (nous utiliserons ici, dans ce domaine, les méthodes de la dynamique moléculaire et du docking), de ”visualiser” l’action d’une drogue sur sa cible. Loin d’être une abstraction, les méthodes numériques permettent ainsi l’accès le plus frontal qui soit aux phénomènes biologiques à l’échelle microscopique.

Enfin, dans la continuité des investissements colossaux effectués dans la génomique, la communauté scientifique se tourne désormais vers le protéome et l’interactome, qui seuls permettront une bonne compréhension des mécanismes fondamentaux du vivant. En particulier, les structures expérimentales disponibles de biomolécules, à travers des bases publiques telles que la Protein Data Bank, croissent exponentiellement. Dans ce contexte, les méthodes numériques sont de plus en plus intéressantes face aux techniques purement expérimentales qui ne sont guère en mesure d’appréhender seules le potentiel d’un tel flux de données.

Parmi les méthodes numériques dont il est question ici, la plupart ont pour origine la chimie théorique. Pour l’expert dans ce domaine, les recherches en biologie et en pharmacologie constituent des champs d’application passionnants, de par la nature des progrès scientifiques qu’ils conditionnent et du fait qu’ils se situent toujours à la limite des progrès technologiques. De nombreux défis, sur les plans théorique aussi bien que technique, dans ce secteur de plus en plus interdisciplinaire, attendent d’être relevés.

Ce manuscrit constitue une étude théorique de l’inhibition de l’activité du domaine SH2 de la protéine Grb2 par la liaison de ligands.

La première partie décrit le contexte de ce travail, qui se situe plus généralement à travers le ciblage de Grb2 SH2 parmi les multiples approches de lutte contre le cancer, de m’étude du vivant à l’échelle moléculaire (domaine interdisciplinaire par excellence), et, plus précisément, de la recherche pharmaceutique.

La seconde partie s’emploiera à détailler les connaissances disponibles sur la cible au moment de débuter ce travail, en incluant les avancées qui sont apparues en parallèle à ce travail.

On trouvera dans la troisième partie un résumé des principaux résultats obtenus au cours de cette thèse. Ceux-ci se classent selon trois approches méthodologiques distinctes : celle de la dynamique moléculaire, du docking et enfin du screening virtuel. On pourra constater que de nouvelles connaissances sur la nature physico-chimique de la liaison de certains ligands sur Grb2 SH2 sont mises en évidence dans le premier cas, tandis que les deux autres approches ont surtout donné naissance à des avancées méthodologiques.

Les publications auxquelles ce travail a donné lieu sont enfin disponibles.

Contexte

Cancer

Avant-propos

Le cancer est une famille de maladies caractérisée par une dérégulation des mécanismes de division cellulaire d'un organisme. [1] La croissance cellulaire incontrôlée qui en résulte parasite le fonctionnement normal de la machinerie biologique, et peut aboutir à des déficiences graves d'un ou plusieurs organes vitaux (principalement reins, foie et poumons). Ainsi, la plupart des cancers humains peuvent causer la mort, et il s'avère que les cancers sont devenus la première cause de mortalité dans les pays développés. [2]

Le nombre de décès en France provoqués par le cancer croît constamment (150000 en 2000, contre 125000 en 1980), parallèlement à l'accroissement de l'espérance de vie (les cancers touchant principalement, et avec une gravité accrue, les personnes les plus âgées). Cette augmentation est toutefois moins importante que celle du nombre de cas diagnostiqués (278000 contre 170000) [3], qui traduit aussi bien les progrès de la médecine anti-cancer[*] que les efforts de santé publique principalement axés sur la prévention et le dépistage.

Le cancer n'est pas seulement un axe crucial de santé publique, c'est également un problème social. [4, 5] Cela peut principalement s'expliquer du fait que la vision populaire du cancer se décline et se déforme bien en dehors de la réalité médicale.

Tout d'abord, au cancer est quasi systématiquement associée l'image de mourants, condamnés pour lesquels la médecine et la science ne peuvent que s'acharner inhumainement et en vain : le cancer représente la mort violente précédée de la déchéance physique, celle que personne ne souhaite et qui inspire universellement la peur. À cela vient s'ajouter le fait que le cancer est bien souvent interprété de façon exclusive soit comme fatalité, soit sous l'angle causal. Dans le premier cas, le cancer est symbole d'injustice et d'incompréhension : il peut frapper n'importe qui, n'importe quand. Dans le second cas, sachant que certaines formes de cancer sont bien connues comme amplifiées par des comportements à risques, une généralisation peut s'opérer : si on en est malade, c'est ”qu'on l'a bien cherché”. Enfin, le cancer est parfois considéré comme une maladie contagieuse… Tous ces éléments combinés font du cancer une maladie honteuse[†] ainsi qu'un facteur important d'exclusion sociale.

En 2002, la lutte contre le cancer est déclarée chantier prioritaire par l'État. Un "plan Cancer" est alors mis en place rapidement sur une durée de cinq ans, et doté d'un budget de 1,5 milliard d'euros. Cependant, le plan Cancer est principalement axé sur des mesures de prévention et de dépistage (effectivement prioritaires sur le court terme au niveau de la santé publique) ; les actions de recherche, centralisées autour de ”cancéropôles” pilotés au niveau ministériel, sont d'une importance toute relative et comprennent au final peu de projets théoriques ou orientés vers l'aspect pharmaceutique. Ainsi, la part de financement des recherches de nouveaux médicaments anti-cancer dans le plan Cancer ne s'élève qu'à 10 millions d'euros fin 2004.

---

[*] Ces progrès sont particulièrement flagrants en ce qui concerne le cancer du sein, qui reste le plus fréquent pour les femmes, et dont le taux de mortalité associé a baissé plus que de moitié de 1980 à 2000.

[†] Maladie honteuse jusque dans la mort ! N'emploie-t-on pas systématiquement la formule pudique "mort des suites d'une longue maladie" sur les faire-part de décès ?

Les organismes de recherche publics tels que l'Inserm[*] ou le CNRS[†], déjà impliqués dans les projets du plan Cancer, participent également de façon indépendante à la lutte contre le cancer, soit en finançant directement des projets de recherche, soit en initiant et participant activement à des projets financés de façon externe. Cette thèse résulte ainsi de la mise au point en 2002 d'une ACI (Action Concertée Incitative) dans la catégorie "Molécules et cibles thérapeutiques" nommée "Nouvelles approches thérapeutiques du cancer : conception et validation de molécules inhibant des interactions protéiques et ciblant la voie Ras induite par les protéines à activité tyrosine kinase".

Fonctionnement du cycle cellulaire et causes du cancer

Les cellules d'un organisme ont une durée de vie limitée. Ainsi, elles ont la capacité de se multiplier afin que l'organisme puisse se maintenir en vie, et il existe de nombreux mécanismes permettant de réguler la population des cellules [6, 7], en favorisant leur prolifération (les gènes codant ainsi pour la division cellulaire sont appelés *proto-oncogènes*), ou au contraire en la freinant (les gènes correspondants sont appelés *suppresseurs de tumeurs*). Une trop faible population de cellules peut provoquer l'activation de mécanismes favorisant l'expression de proto-oncogènes, et à l'inverse, la détection de mutations pourra favoriser les suppresseurs de tumeurs, ce qui permettra de procéder à la réparation ou éventuellement à l'élimination des cellules endommagées avant que celles-ci ne puissent mettre en danger l'organisme en se multipliant.

En temps normal, l'ensemble de ces mécanismes permet de maintenir un équilibre entre le nombre de cellules résultantes de la division cellulaire et le nombre de cellules mortes, tout en régulant d'éventuelles altérations du code génétique. Cela permet aussi également de réparer les pertes cellulaires accidentelles ponctuelles dues par exemple à des blessures, assurant ainsi le bon fonctionnement des organes et des tissus. Lorsque cet équilibre est rompu, cela peut avoir pour effet une multiplication anarchique de certaines cellules, lesquelles peuvent se concentrer, formant des *tumeurs*, lesquelles caractérisent le cancer lorsque leur taille ne permet plus leur retrait par des interventions médicales simples. Dans de nombreux cas, ces tumeurs peuvent menacer directement le bon fonctionnement d'un ou plusieurs organes vitaux ou du système sanguin, ou sont susceptibles de le faire en migrant dans l'organisme (on parle alors de métastases). Le cancer met alors gravement en danger la vie de l'individu affecté et aboutit souvent à sa mort en l'absence de traitement. À un tel stade, le cancer met alors gravement en danger la vie de l'individu affecté et aboutit souvent à sa mort en l'absence de traitement (90% des cas de mortalité dues au cancer sont consécutives à la formation de métastases [8]).

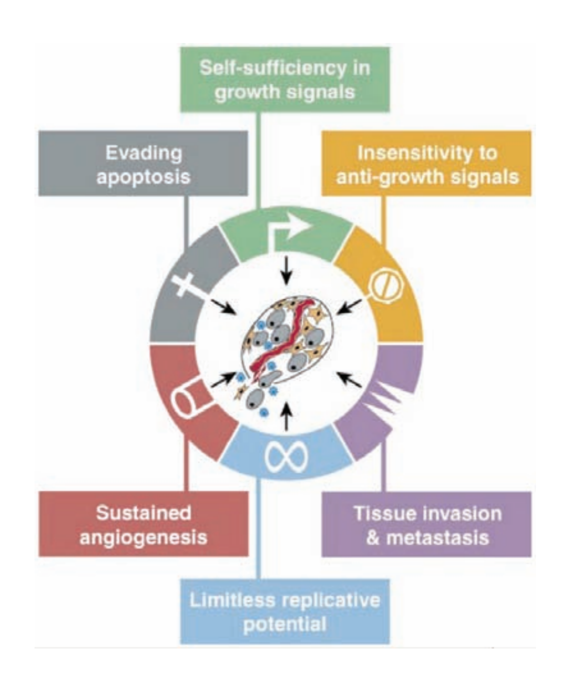
Quelle qu'en soit la cause, le développement d'un cancer a toujours pour origine la mutation d'un ou plusieurs proto-oncogènes, qui sont alors appelés *oncogènes* [9] : le cancer est une maladie génétique. La totalité des cancers résultent une série concertée de mutations génétiques et non d'une seule [10-13], ce qui explique l'augmentation du risque avec l'âge et le long cycle de développement de la maladie que l'on observe généralement. Le développement du cancer est souvent lié à une surexpression de protéines exprimées par les proto-oncogènes mutés en oncogènes, donc impliquées - lorsque leur concentration dans l'organisme est appropriée - dans les processus normaux de croissance cellulaire, et aboutissant à une multiplication incontrôlée des cellules affectées. Le cancer peut également être favorisé par la mutation d'un ou plusieurs gènes suppresseurs de tumeurs [14, 15], aboutissant à la diminution ou la disparition de biomolécules chargées de réparer ou d'éliminer les cellules endommagées. La plupart des cancers combinent

[*] http://www.inserm.fr/fr/questionsdesante/dossiers/cancer/
http://www.inserm.fr/fr/recherches/etats_des_lieux/att00002003/CANCER2002_v3.4-num.pdf

[†] www.cnrs.fr/SDV/cnrscancer.html

ces deux causes, une mutation pouvant en provoquer d'autres par cascade, ceci étant accéléré à chaque fois qu'un gène suppresseur de tumeurs est touché. Cela aboutit à la formation d'au moins une cellule cancéreuse, pour laquelle les signaux génétiques qu'elle contient et favorisant sa prolifération prévalent, suite à un nombre critique de mutations (un nombre trop faible de mutations étant régulé naturellement), sur ceux codant pour sa régulation.

Les différentes mutations nécessaires à la formation d'une cellule cancéreuse sont à présent bien caractérisées. [13] Les mutations favorisant la formation d'une cellule cancéreuse confèrent à celle-ci une autonomie par rapport à l'organisme hôte : la cellule cancéreuse génère ses propres facteurs de croissance [16] (ce qui dérégule les voies de signalisation correspondantes [17]), et est insensible aux mécanismes de régulation de la population cellulaire par l'organisme, en particulier la mort cellulaire programmée [18, 19]. D'autre part, la cellule cancéreuse acquiert une grande capacité de multiplication [20, 21] et de prolifération dans l'organisme, ainsi qu'une meilleure résistance aussi bien au système immunitaire qu'à un traitement extérieur (cela se traduit, en particulier, par la mutation du gène suppresseur de tumeur p53 [22]). Une cellule cancéreuse est une cellule pour laquelle l'ensemble de telles mutations rend irréversible sa mutation infinie dans l'organisme, à moins qu'un traitement approprié ne soit effectué sur la personne affectée.



Il est encore impossible de déterminer l'origine d'un cancer donné antérieurement à la série de mutations génétiques qui le provoque ; on estime même souvent que ces mutations peuvent survenir spontanément. Toutefois, il s'agit d'évènements anormaux et étant donné que seule une combinaison de plusieurs types de mutations provoque le cancer, on estime qu'un des préalables à la survenue du cancer est une instabilité du génome sur un intervalle de temps important. Il a été démontré que certains facteurs extérieurs précis peuvent en être la cause. On peut citer l'exposition répétée à des substances dites *cancérigènes* [23, 24], qui peuvent être des drogues (alcool, tabac...), des produits chimiques (solvants, pesticides...) ou bien des produits naturels (amiante, matériaux radioactifs...). De nombreux cancérigènes ont été - et sont toujours, pour certains d'entre eux - massivement employés dans l'industrie (arsenic, amiante, nickel, goudrons...), si bien que la part des expositions professionnelles parmi les causes du cancer n'est certainement pas négligeable.

Certains cancérigènes sont caractéristiques d'un type précis de cancer, par exemple le tabac dont la consommation est parfaitement corrélée démographiquement au nombre de cas de cancer du poumon. La combinaison de plusieurs substances cancérigènes augmente encore les risques ; en particulier, l'association alcool / tabac est connue pour son incidence très importante sur les cancers de la bouche, de la gorge et de la vessie. On estime, plus généralement, qu'une mauvaise hygiène de vie (un régime alimentaire trop riche en graisses, par exemple), ou bien un environnement pollué favorisent l'apparition de cancers. Des mutations génétiques prédisposant à certaines formes de cancer peuvent également être héritées [25] (c'est le cas d'un français sur 60), par exemple les mutations du gène BRCA1 reliées aux cancers

du sein et des ovaires. Cet aspect semble cependant avoir moins d'incidence que les facteurs environnementaux [26]. Enfin, de façon moins fréquente, certains virus, contenant dans leur génome des oncogènes ou gènes inactivant les suppresseurs de tumeur, ou bien favorisant indirectement des proto-oncogènes de leur cellule hôte lors de leur réplication, peuvent provoquer des mutations pouvant aboutir à la formation de tumeurs cancéreuses [27, 28].

Les différentes approches de la lutte contre le cancer

La lutte contre le cancer est constituée de multiples approches souvent complémentaires. Il en va de même en ce qui concerne les aspects aussi bien médicaux que scientifiques. Dans un tel contexte, il est évident que la lutte contre les différentes formes de cancer ne pourra pas être le fait d'une seule technique scientifique, d'un seul axe de recherche. En énumérant brièvement les principales directions qui peuvent être suivies, nous pouvons donner l'impression qu'il s'agit de domaines distincts ; bien au contraire, dans le cadre le la recherche anti-cancer l'interdisciplinarité est d'une importance cruciale, tant le sujet d'étude est complexe.

En premier lieu, la prévention est fondamentale car il est établi que plus un cancer est décelé précocement et les traitements thérapeutiques effectués en conséquence, meilleures sont les chances de survie du patient. La prévention inclut des actions de dépistage et de sensibilisation du public, mais aussi des études cliniques et épistémologiques. Ces dernières permettent cerner avec plus de précision les facteurs de risque, ce qui constitue un préalable indispensable à toute action de prévention visant les populations à risque, mais permet également de recouper les connaissances issues des recherches en biologie, et de fournir au milieu médical des données statistiques précieuses pour la prise en charge et le suivi des patients.

Si des examens réguliers sont indispensables pour les formes de cancer les plus courantes, on peut également isoler les facteurs de risques potentiels chez un individu. Une fois le cancer diagnostiqué, le choix d'une ou plusieurs approches dépend de la nature du cancer, de son avancement, ainsi que de l'état de santé général du patient concerné. L'intervention chirurgicale [29] (extraction de tumeur ou amputation de l'organe affecté), plus ancienne méthode thérapeutique de traitement contre le cancer, est toujours largement pratiquée dès lors que la tumeur est localisée et solide. Afin d'éviter sa résurgence après l'opération, on utilise la *radiothérapie* [30] et la *chimiothérapie* [31], qui peuvent d'autre part s'avérer suffisantes lorsque la tumeur n'a pas atteint une taille critique, et peuvent constituer au minimum des traitements palliatifs efficaces dans les cas où le développement et la nature de la tumeur ne permettent plus un traitement curatif. Toutefois, ces deux méthodes sont connues pour provoquer nombre d'effets secondaires désagréables car elles ne ciblent pas exclusivement les cellules cancéreuses. D'autres méthodes thérapeutiques telles que l'*hormonothérapie* [32, 33] sont accessibles, ne ciblant plus les cellules cancéreuses mais favorisant ou non des biomolécules (protéines, anticorps, hormones...) qui leurs sont associées, cependant elles n'ont souvent encore qu'un rôle complémentaire du fait de leur efficacité plus modeste ou de leur domaine d'action plus limité. Ainsi, il est courant d'employer simultanément plusieurs techniques parmi celles que nous venons d'évoquer. [34]

Parallèlement à cela, de nouvelles thérapies sont en développement. Ainsi, de nombreux travaux de recherche portent sur la *thérapie génique* [35], consistant à insérer dans les cellules des gènes à fonctions thérapeutiques. Une des grandes difficultés de cette approche réside dans la mise au point de vecteurs efficaces (virus, bactéries...) permettant de transférer l'information génétique dans les cellules, tandis qu'ils sont ciblés par le système immunitaire de l'individu. Le développement de vaccins anti-cancer ou *immunothérapie* [36-40] dont le but est aussi bien de prévenir l'infection de l'organisme par des agents cancérigènes que de favoriser les réponses immunitaires au développement d'une tumeur ultérieure potentielle, fait également l'objet d'efforts de recherche poussés.

Un autre domaine de recherche prometteur de la lutte anti-cancer, et dans lequel nous nous situons dans le cadre de cette étude, consiste à mettre au point des inhibiteurs chimiques plus spécifiques [41, 42], avec potentiellement une efficacité accrue et des effets secondaires [43] moins importants pour le patient. [44] De tels médicaments s'avèrent déjà utiles en complément de traitements anti-cancer plus conventionnels. [45] Ces inhibiteurs peuvent être des molécules en partie ou en totalité synthétiques, et ne ciblent ni le système immunitaire (immunothérapie), ni certaines hormones (hormonothérapie), mais des protéines [46, 47] et des voies de signalisation [48-51] dont le rôle dans le développement du cancer est bien caractérisé. Cela nécessite préalablement une connaissance plus poussée des mécanismes biomoléculaires correspondants. [52] À ce niveau, l'interprétation de résultats de recherches en génétique peut s'avérer précieux. [53]

Quelques liens utiles sur le cancer :

Ligue contre le cancer : http://www.ligue-cancer.asso.fr

Fédération Nationale des Centres de Lutte Contre le Cancer : http://www.fnclcc.fr/

Rapport de la commission d'orientation sur le cancer (2002) : http://www.sante.gouv.fr/htm/dossiers/cancer/

Mission interministérielle pour la lutte contre le cancer : http://www.plancancer.fr/

Cancer Medicine 5th edition (BC Decker) :
http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowTOC&rid=cmed.TOC&depth=10

American Cancer Society : http://www.americancancersociety.org

National Comprehensive Cancer Network : http://www.nccn.org

Résumé de l'ACI qui a permis de financer cette thèse :

Le but de ce projet est de concevoir des agents pharmacologiques ciblés, capables d'inhiber des voies de signalisation dérégulées par la présence de protéines codées par des oncogènes et conduisant à la formation de cancers.

L'approche choisie est nouvelle mais donne déjà des résultats. Elle consiste à inhiber les interactions entre ces protéines et leurs effecteurs ou entre des protéines situées en aval dans la voie dérégulée. Les cibles sélectionnées sont situées dans la voie de signalisation Ras-dépendante impliquée dans des cancers à la fois nombreux et souvent fréquents : ERBB2 (sein, ovaire), Grb2 (sein, ovaire et certaines leucémies comme la LMC), RasGAP dans les cancers exprimant Ras oncogénique (30% des tumeurs, 90% pancréatiques).

La réunion de biologistes, pharmacologues, chimistes et physicochimistes académiques et du secteur industriel désireux de mettre en commun leurs compétences va permettre, dans un premier temps de concevoir des agents peptidiques compétiteurs des interactions ciblées, et de les valider par des tests cellulaires et vivo en les vectorisant, puis de procéder à des études structurales des complexes entre protéine et peptide inhibiteur afin de définir les points d'interaction pour concevoir grâce à l'aide de la modélisation moléculaire « in silico », des molécules non peptidiques selon une approche de chimie combinatoire « ciblée ».

L’étude théorique du vivant à l’échelle moléculaire

La lutte contre le cancer comme objectif de la recherche scientifique s’inscrit dans le cadre plus général des sciences du vivant. On entend par là un ensemble de disciplines scientifiques, telles que la biochimie et la génétique, issues des progrès réalisés au XXème siècle autour de la biologie, et dont les applications fondent désormais une très grande part des découvertes en pharmacologie et en médecine. Les sciences du vivant reposent avant tout, au niveau théorique, sur la connaissance du fonctionnement d’un organisme vivant. Une des approches possibles pour décrire les mécanismes correspondants, et sur laquelle ce travail repose, est celle de la biochimie, pour laquelle les mécanismes de base sont des interactions entre *biomolécules*. L’étude théorique de ces biomolécules est rendue possible par le progrès de certaines techniques expérimentales, et surtout grâce à la banalisation des calculs informatiques ; il s’agit d’un axe de recherche relativement récent.

Dans cette partie, nous décrirons tout d’abord les deux classes fondamentales de biomolécules : les acides nucléiques (ADN et ARN) et les protéines. Les premières contiennent le ”mode d’emploi” de l’organisme, tandis que les secondes œuvrent à son fonctionnement en assurant la communication entre les cellules. Nous expliquerons ensuite en quoi le champ de vision des sciences du vivant s’est considérablement élargi ces dernières années, entre autres grâce au poids croissant des méthodes de calcul informatique dans la recherche scientifique, passant d’une vision centrée sur le *génome* à une vision plus large qui inclut l’ensemble des protéines, ou *protéome*, puis ensuite l’ensemble des interactions entre biomolécules, ou *interactome*.

Structure des biomolécules

Des acides nucléiques à l’information génétique

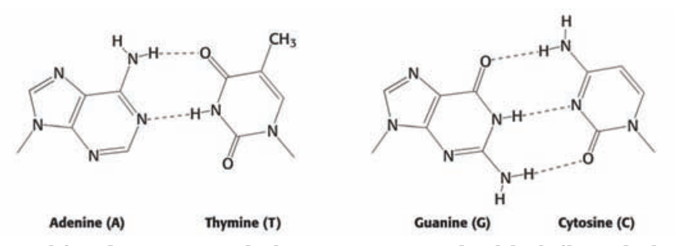
Les acides nucléiques s’observent sous deux formes polymériques : l’acide ribonucléique (ARN) et l’acide désoxyribonucléique (ADN), dans lesquelles des groupes fonctionnels appelés *nucléotides* sont greffés sur un squelette polymère pentose-phosphate. Quatre nucléotides différents sont rencontrés dans la structure de l’ADN* : adénine (A), thymine (T), cytosine (C) et guanine (G).

Le squelette pentose-phosphate de l’ADN et les 4 bases possibles correspondantes

* Dans l’ARN les thymines sont mutées en uraciles, et les déoxyriboses du squelette en riboses.

ADN et ARN ont des fonctions différentes de par leur structure. La structure de l'ADN fut déterminée par Watson, Crick et Franklin en 1953 [54-56], et a pour particularité le positionnement des nucléotides à l'intérieur d'une double hélice formée par deux chaînes pentose-phosphate antisymétriques. Ce positionnement dans un espace si confiné est rendu possible par un appariement des nucléotides par paires : adénine et thymine d'une part, cytosine et guanine d'autre part, sont reliées d'une hélice à l'autre par des liaisons hydrogène. L'information génétique est ainsi ”doublée” et stabilisée : le rôle biologique de conservation tenu par l'ADN apparaît alors évident. L'ARN a une structure en simple hélice dans laquelle les nucléotides sont exposés, permettant la réplication et l'expression directe de la séquence.





Appariement des nucléotides au sein de la structure en double hélice de l'ADN

Du génome au protéome

Les protéines sont également des polymères, constitués à partir des 20 acides aminés naturels. La signification du codage génétique donne le lien entre les séquences nucléiques et peptidiques : *les séquences génétiques codent les séquences protéiques* [57] ; par séquences de trois nucléotides de l'ARN ou *codon* (voir ci-contre) [58, 59]. La synthèse d'une protéine a plus précisément pour origine la transcription d'une partie spécifique fonctionnelle de l'ADN (un gène) sous forme d'ARN*. [60] Le mécanisme de transcription de l'ARN vers des protéines fait intervenir le *ribosome*, ensemble d'une centaine de protéines enzymatiques qui sont les ouvrières de cette ”chaîne de production”. [61]

First position (5′ end)	Second position				Third position (3′ end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Au niveau du protéome, un niveau de complexité supplémentaire† est présent par rapport au génome : alors que l'ADN et l'ARN possèdent des structures invariables, l'activité et la fonction des protéines dépendent intrinsèquement de leur géométries‡, qui définissent un large espace conformationnel.§

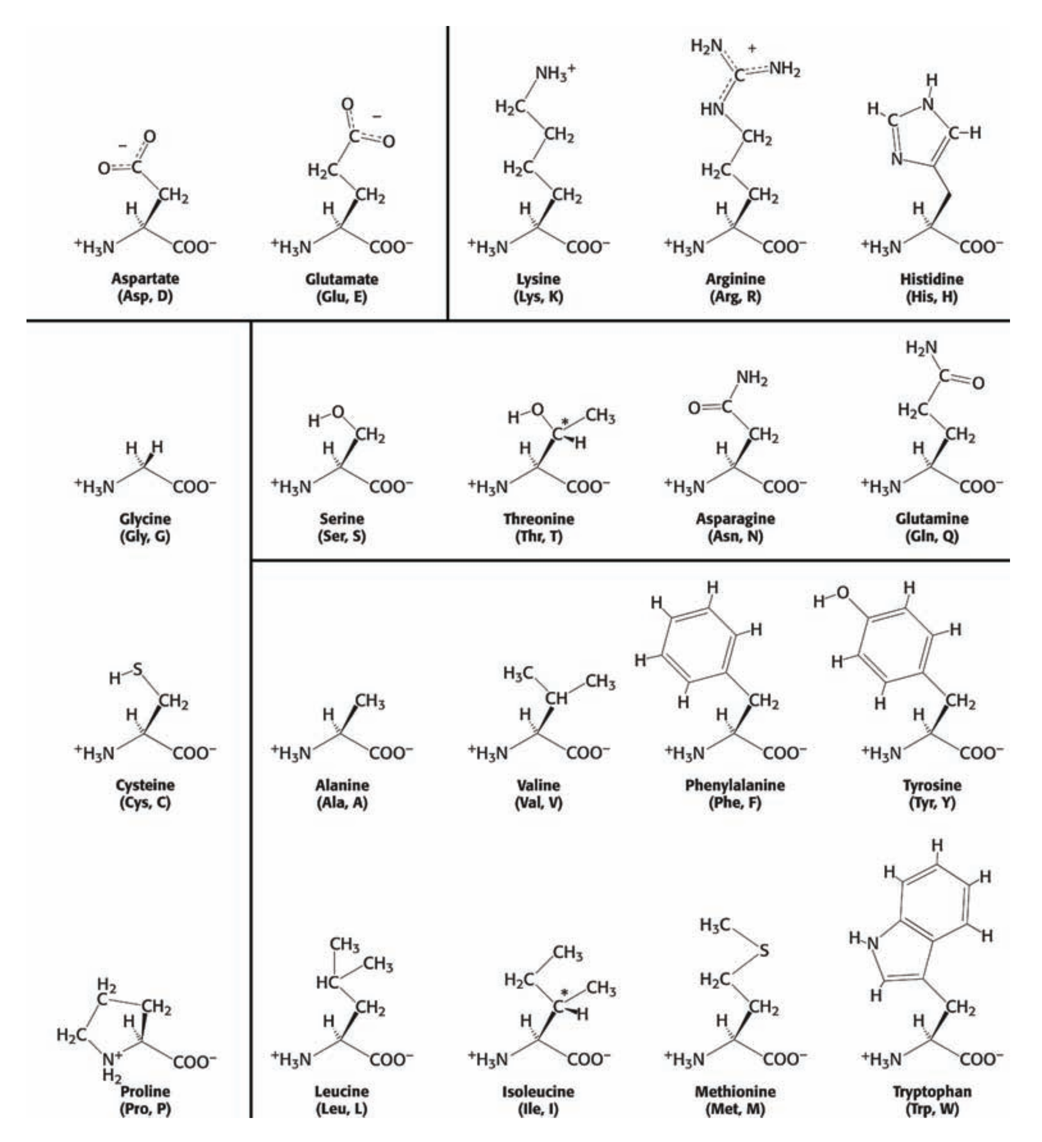
---

* On appelle ce processus *transcription de l'ADN* ; proche sur de nombreux points du processus de réplication de l'ADN intervenant lors de la division cellulaire. Certaines protéines dites *régulatrices de gènes* servent à déterminer quelles parties de l'ADN doivent être transcrites et quelles parties sont ignorées, définissant ainsi les segments sur les chromosomes correspondant aux gènes.

† On peut noter qu'alors que dans l'organisme humain la taille moyenne des protéines est de 300 résidus, on est très loin d'observer 20^{300} soit environ 10^{390} séquences peptidiques possibles. L'ordre de grandeur des séquences significatives est estimé à 10^5.

‡ De façon remarquable, l'adoption par une protéine d'une conformation ”anormale” peut changer sa fonction biologique et aboutir à des pathologies particulièrement graves. C'est le cas de la ”maladie de la vache folle”, dans laquelle une protéine du cerveau (un prion) subit un changement conformationnel se propageant aux autres prions, formant des aggrégats à la source d'un processus dégénératif mortel.

§ La conformation d'une protéine peut être caractérisée par l'orientation relative des différentes unités du squelette. On distingue pour ce faire les hélices α (caractérisées par des liaisons H intra-chaîne, et représentées graphiquement sous la formes de cylindres) des feuillets ß (correspondent à l'association parallèle ou anti-parallèle de deux chaînes distinctes par des liaisons H inter-chaînes, par des flèches plates). On appelle tour (ß ou Ω lorsque l'angle est très marqué) le passage d'une forme géométrique à l'autre.



Liste des 20 acides aminés classés naturels classés par propriétés,respectivement : acides (D, E), basiques (K, R, H), structure spécifique (G, C, P), hydrophiles (S, T, N, Q) et hydrophobes (A, V, F, Y, L, I, M, W)

Il a été démontré que la séquence d'une protéine détermine la structure géométrique globulaire adoptée par un processus de repliement sur elle-même qui survient après sa synthèse. Les principes de base de ce mécanisme sont connus [62], ainsi que des résultats statistiques*. Sa prédiction, désignée comme le *problème de repliement des protéines* [63], a un intérêt évident, mais la tâche s'avère si complexe qu'elle en constitue une sorte de "saint Graal" de la biochimie.

Ainsi, alors que l'étude du *génome* correspond fondamentalement à l'exploration d'un espace séquentiel (sur quatre bases), celui du *protéome*, en plus de l'extension d'une partie de cet espace (cette fois codé sur 20 bases), lui ajoute un espace conformationnel† beaucoup plus vaste. [64] L'exploration de ce protéome est très délicate, et ce même si on ne s'intéresse qu'à une protéine bien définie, à moins qu'une structure expérimentale de celle-ci soit disponible, ce qui fut le cas pour ce travail. Les techniques de la modélisation moléculaire peuvent alors être mises en œuvre.

---

* Par exemple, les résidus Met et Glu se retrouvent plus fréquemment dans des hélices α, Val et Ile dans des feuillets ß, Gly et Pro dans les tours. L'Arg est le seul acide aminé qui ne semble pas avoir de "préférence" particulière.

† On passe alors de la *structure primaire* (linéaire) des protéines qui se limite à la séquence à la *structure secondaire* (globulaire). On parle de *structure tertiaire* lorsque l'on considère en plus les interactions "longue distance" sur les chaînes peptidiques qui entre autres différencient structuralement les résidus hydrophiles des résidus hydrophobes. Enfin, la *structure quaternaire* se réfère à l'organisation de chaînes peptidiques analogues en super-structures.

Du génome à l’interactome : description du fonctionnement d’un organisme vivant

Du protéome à l’interactome

Les interactions entre protéines peuvent être identifiées, regroupées au sein de *voies de signalisation*, et représentées schématiquement sous la forme de graphes d’interactions. [65, 66].

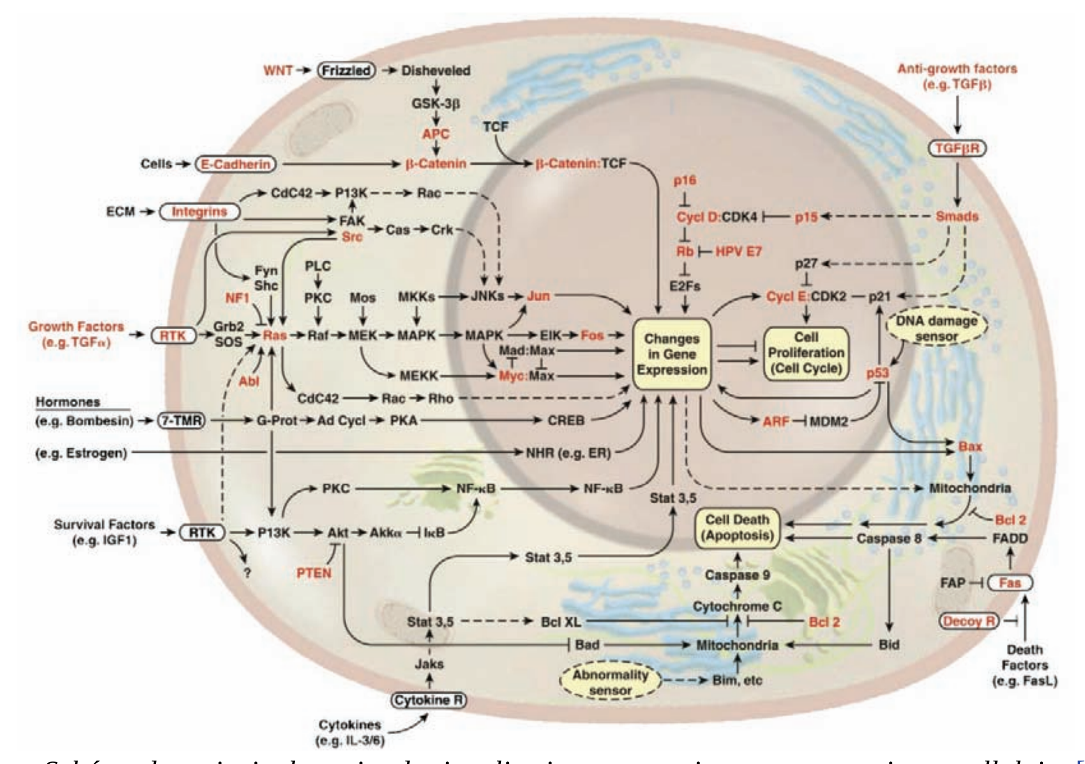


Schéma des principales voies de signalisation connues intervenant au niveau cellulaire [13]

L’étude sur le plan moléculaire du fonctionnement d’un organisme vivant peut ainsi s’effectuer à différents niveaux conceptuels successifs. **Le *génome* repose sur l’espace des séquences de nucléotides, le *protéome* y ajoute l’espace géométrique des protéines correspondantes, et l’*interactome* la liste et la nature des interactions possibles qui en découlent**. Le génome humain est à présent décrypté, et concernant les protéines le nombre de structures expérimentales résolues et accessibles [67, 68] croit exponentiellement.* La cartographie de l’interactome n’en est quant à elle qu’à ses balbutiements.†

Dans le cadre de ce travail, nous étudions une cible protéique précise et en particulier l’inhibition de son activité au sein d’une voie de signalisation correctement caractérisée. Il ne nous apparaît pas important dans ces conditions de détailler l’état actuel des recherches sur l’interactome, qui peuvent inclure des aspects autres que purement géométriques et chimiques, tels les effets cinétiques.‡

---

* Au 23 août 2005, la base PDB (Protein Data Bank) recence 29532 structures protéiques. La portion de ces structures qui correspondent à des complexes, et qui sont donc d’un grand intérêt afin de relier protéome et interactome au niveau de l’espace géométrique et non plus seulement au niveau de l’espace des interactions, n’est toutefois que de 1316 (4,4%).

† Si l’interactome est un réseau routier, on n’en connaît pour le moment qu’une partie des autoroutes sous la forme de voies de signalisation. La plupart des interactions entre biomolécules sont inconnues à l’heure actuelle.

‡ En effet, plusieurs interactions peuvent être ”en concurrence” au sein du même milieu extra- ou intra-cellulaire. Ces considérations peuvent être cruciales pour des applications d’ordre médical.

Constitution moléculaire d’un organisme vivant

La ”composition chimique” d’un organisme vivant ne saurait bien évidemment se réduire aux acides nucléiques et aux protéines. Toutefois, ces deux classes fondamentales de biomolécules sont à la base du fonctionnement des organismes et suffisent à caractériser le cadre de cette thèse. De façon générale, si l’on exclut les biomolécules de taille importante (acides nucléiques, protéines, membranes…), la diversité en terme chimiques d’un organisme vivant est limitée – les organismes vivants les plus simples sont ainsi construits autour d’une centaine de petites molécules organiques [69] – en comparaison du nombre de molécules organiques théoriquement synthétisables (estimé à 10^{60} [70]). ”L’espace biologique” est de nombreux ordres de grandeur plus petit que ”l’espace chimique”. [71]

Un aspect important de la nature d’un organisme vivant observée sous l’angle de sa constitution chimique doit néanmoins être mentionné. Les systèmes en biochimie ne doivent pas être considérés comme des systèmes chimiques classiques dans lesquels les molécules sont en interaction en milieu dilué. Le fait que le corps humain soit constitué d’environ 70% d’eau signifie en particulier du point de vue chimique que la plupart des biomolécules se trouvent en solution à des concentrations particulièrement élevées d’environ 30% [71], formant une ”masse” en mouvement et en interaction. [72]

L’ère de la post-génomique

Conceptuellement, le code génétique d’un organisme vivant correspond à son ”programme de fonctionnement” ; ainsi, le déchiffrage de ce code constitua pour la communauté scientifique un objectif fondamental. Étant donné que biologiquement toute fonction découle de ce codage, le but ultime de la génétique fut l’interprétation de toute fonction biologique à partir d’un code génétique.

Cet enthousiasme est à présent retombé. La génétique s’avéra un outil remarquable sur le plan médical et a permis d’identifier la source de nombreuses pathologies, tout en fournissant de précieuses indications sur des traitements pharmaceutiques potentiels. Mais il est à présent évident que l’étude du génome ne peut pas être la ”recette miracle” qui révolutionnera la compréhension du vivant. En particulier, l’hypothèse réductionniste ”un gène = une fonction biologique” s’avéra fausse. Pour un grand nombre de gènes, la fonction biologique correspondante est inconnue et il n’est même pas possible de déterminer s’il y en a une.

Pour ces raisons, on estime souvent que l’effort investi dans la génétique doit à présent être étendu en direction de l’interactome. En effet, si la génétique permet d’identifier de nombreuses pathologies, elle reste souvent impuissante dès lors qu’il s’agit de les corriger. Si la thérapie génique est un domaine de recherche plus connu, la mise au point de médicaments ciblant l’interactome est également un axe prometteur en particulier sur le plan pharmaceutique [73]. Le travail de cette thèse se situe dans ce dernier domaine.

Importance des simulations numériques

L'informatique : une révolution scientifique

Dans le passé, les modèles physiques théoriques de la matière ne pouvaient être vérifiés que dans des circonstances expérimentales particulières, dans lesquelles des variables physiques isolées peuvent être mesurées. D'une façon générale, la mise au point de modèles théoriques par le scientifique résulte classiquement d'une approche *réductionniste* de la nature[*]. Le point faible d'une telle démarche est que les modèles validés obtenus seront par nature *réducteurs* : la complexité n'est pas accessible.

L'informatique a été une révolution scientifique, car elle a permis à la complexité d'être – en partie – accessible scientifiquement. La possibilité de conduire des calculs étendus et automatisés permet l'application un modèle ou une combinaison de modèles à des systèmes dont la taille et la complexité ne les rend pas traitables par d'autres moyens. Les techniques informatiques permettent ainsi de se rapprocher de la réalité physique, validant ou invalidant les modèles avec plus de rigueur. Si l'informatique n'est pas à l'origine de la découverte de théories scientifiques, bien souvent elle seule permet d'étendre le champ d'application des modèles qui en découlent.[†]

Les simulations numériques : une nouvelle méthode de recherche

L'apparition de moyens de calcul informatique de plus en plus importants (la loi de Moore, jusqu'ici vérifiée, postule que la puissance de calcul des microprocesseurs double tous les 18 mois) à partir des années 1950 a altéré le binôme théorie-expérience de la recherche scientifique en insérant une méthode intermédiaire : la simulation numérique. Fondamentalement, **aussi bien l'aspect théorique que l'aspect expérimental de la recherche sont intégrés dans l'outil informatique**, le premier à travers des programmes et des algorithmes, et le second à travers des simulations et des données. Ainsi, la conduite d'une simulation peut être analogue, selon les circonstances, à la celle d'une expérience ou à la mise en place d'une théorie, rendant l'une comme l'autre plus accessibles.

Un autre avantage considérable de l'informatique doit également être mentionné. **Une simulation numérique n'est soumise à aucune autre contrainte matérielle que celle du temps de calcul nécessaire.** Le choix entre mesure expérimentale et simulation numérique[‡] n'a ainsi pas lieu d'être lorsque l'expérience est trop contraignante (par exemple, elle requiert des conditions de température et de pression délicates, elle peut s'avérer dangereuse…), voire impossible. Les simulations numériques peuvent aller encore plus loin, en permettant de repousser artificiellement certaines contraintes physiques[§], afin de mieux appréhender le comportement d'un système.

---

[*] L'approche réductionniste consiste à fractionner un système complexe en sous-systèmes plus simples, jusqu'à ce que tous les sous-systèmes puissent être définis de façon satisfaisante par des modèles théoriques. Pour valider un sous-système donné pour lequel des hypothèses théoriques sont avancées il faut mettre en place une expérience dans laquelle interviennent le moins possible de phénomènes physiques externes au sous-système décrit. Cette contrainte est souvent particulièrement difficile à satisfaire.

[†] Les premiers modèles physiques des gaz parfaits ne pouvaient être validés que par des mesures sur des gaz rares, et les modèles de cristaux devaient l'être avec des échantillons très purs. S'ils permettent bien d'appréhender les lois fondamentales qui leur sont associées, ils ne permettent aucunement de décrire un gaz ou un cristal quelconque. Et en chimie théorique, les méthodes *ab-initio* se limitent à des assemblages simplistes en l'absence de moyens numériques.

[‡] Remarquons que de nombreux appareillages modernes incorporent, afin d'affiner ou d'interpréter les mesures (par exemple, des données spectrométriques), des traitements par simulations numériques. L'outil informatique n'est donc pas seulement indispensable au niveau conceptuel, mais également au niveau technique, même si ce dernier point n'est pas aussi "visible".

[§] Dans le cas de systèmes moléculaires, on peut franchir des barrières de potentiel afin d'explorer au mieux un espace conformationnel. De plus, la mise en place de transformations alchémiques est à la base des méthodes de calcul d'énergie libre.

Pour de nombreux systèmes, la mise en œuvre de la simulation correspond à la mise en place d'un modèle et d'un protocole qui constituent un compromis acceptable entre le réalisme des résultats et l'accessibilité en termes de temps de calcul. Nous détaillerons à quel niveau des approximations devront être effectuées dans le cadre de la simulation de systèmes biomoléculaires.[*]

La simulation de systèmes biologiques par modélisation moléculaire

Les différentes méthodes que nous avons utilisées pour ce travail appartiennent à la classe de la modélisation moléculaire. Nous les avons appliquées à l'étude de systèmes biomoléculaires, en général à la fois complexes chimiquement et difficilement accessibles expérimentalement ; les moyens informatiques sont dans un tel contexte indispensables.

Le champ d'application des techniques de modélisation moléculaire est immense, à l'image du spectre des connaissances et des domaines scientifiques qui sont susceptibles d'intervenir. Nous nous limiterons aux systèmes que nous nous apprêtons à étudier, de type biologique.

La modélisation moléculaire peut être précieuse dans l'identification de cibles thérapeutiques, typiquement des récepteurs protéiques dont le rôle spécifique dans une pathologie est caractérisé. Elle peut s'avérer ensuite indispensable lors de la mise au point d'inhibiteurs chimiques pour cette cible. Un bon exemple du rôle positif de telles méthodes dites de *structure-based drug design* [74] est constitué par la mise au point d'inhibiteurs de l'HIV protéase [75]. Des techniques telles que la modélisation par homologie peuvent permettre de pallier à l'absence de structures expérimentales complètes dans le cas de certaines cibles d'intérêt majeur, telles que les récepteurs de type GPCR. [76, 77] De nombreuses approches sont disponibles et peuvent être combinées afin de produire des résultats scientifiques de grand intérêt, en particulier sur le plan médical. [78]

La modélisation peut, plus généralement, profiter directement des progrès dans une grande variété de domaines. En premier lieu, on peut citer les progrès de l'informatique, aussi bien sur le plan matériel (puissance de calcul brute, architectures parallèles et distribuées) que logiciel (outils spécialisés, techniques de programmation). La disponibilité plus importante de structures expérimentales est également très profitable ; les progrès technologiques des appareillages ont un impact positif pour les modélisateurs. Il en va de même pour les avancées purement théoriques, ainsi que les nouvelles découvertes en biologie et biochimie, qui peuvent permettre d'améliorer les modèles au niveau conceptuel. Nous nous situons donc, avec ce travail, dans un contexte résolument interdisciplinaire.

Relier conceptuellement des domaines si différents constitue bien évidemment un défi scientifique particulièrement intéressant, mais la médaille a son revers. En effet, nous allons être confrontés rapidement à une grande hétérogénéité au niveau des moyens mis en œuvre, des appareillages, des concepts, des données… À notre niveau, on peut supposer que les difficultés qui vont apparaître seront plus d'ordre technique, au niveau de la mise en œuvre des moyens informatiques, que purement conceptuels.

[*] Dans le domaine de la modélisation moléculaire, les conditions permettant d'obtenir un modèle très réaliste sont connues : il suffit d'appliquer les lois de la mécanique quantique. Mais en contrepartie, les calculs seront si complexes que l'accessibilité du système sera la plupart du temps impossible à atteindre, limitant les simulations à des modèles n'impliquant qu'un petit nombre d'atomes ($<<$ 1000). La simulation de systèmes de grande taille tels que les biomolécules nécessite de rabaisser le réalisme et la représentativité des simulations. En particulier, les effets électroniques seront particulièrement difficiles à caractériser.

Drug design

Étapes de la mise au point d’un médicament

Identification de la cible pharmaceutique

L’identification d’une cible pharmaceutique peut se faire par différents moyens, parmi lesquels viennent en premier lieu les méthodes de la biologie cellulaire et moléculaire. Il faut leur ajouter des techniques plus récentes issues de la génomique et de la bioinformatique. Une fois la cible identifiée, il convient de rassembler scrupuleusement toutes les informations accessibles à son sujet, à commencer par les informations bibliographiques. Une cible peut enfin être caractérisée de différentes façons : une structure, un mécanisme biologique, une séquence nucléique ou protéique, *etc.* Chacune de ces caractérisations ouvre des possibilités d’étude dans l’optique de la découverte de médicaments actifs.

On peut énumérer plusieurs exemples d’informations utiles, à différents niveaux. Ainsi, il est essentiel de disposer d’une structure fiable de la cible afin d’accéder aux techniques dites de *structure-based drug design* dont cette thèse se fait l’écho.[*] D’autres aspects sont particulièrement importants. Quelle est l’accessibilité expérimentale de la cible, aussi bien *in vitro* que *in vivo*[†] ? Dans les deux cas, il est utile de disposer de protocoles de mesures biologiques de l’activité d’une molécule donnée, qui soient les plus fiables et accessibles qu’il se peut : si le coût d’un tel test est important, cela limitera considérablement le nombre de molécules que l’on pourra envisager de tester expérimentalement par la suite. Et, avant tout, toujours dans l’optique du *drug design*, est-ce que la cible peut être accessible dans les conditions physiologiques par une molécule administrée de façon externe ?

Dans tous les cas, **toute information initiale sur la cible est susceptible de fournir des moyens d’adapter les protocoles de recherche de médicaments qui s’ensuivront**.[‡] L’étude de l’interactome (validation d’une cible et étude sa fonction biologique à travers l’identification du mode d'action des protéines associées) vient compléter dans ce contexte l'étude du génome (identification d’une cible à partir de l’expression des gènes codant pour les protéines impliquées).

Identification des composés prometteurs (hits)

Une fois la cible pharmaceutique identifiée, il faut soit tester un ensemble de molécules-candidates sur cette cible, selon un processus qualifié de *screening*[§] (criblage), soit procéder à la mise au point rationnelle de molécules, soit opter pour une combinaison de ces deux approches de base.

---

[*] Dans l’idéal, des structures expérimentales résolues par RMN ou RX seront disponibles ; si tel n’est pas le cas, il faudra construire un modèle et être particulièrement vigilant lors de sa validation. En effet, si le modèle de départ n’est pas correct, le temps qui sera ensuite passé dans les étapes suivantes de recherche le sera en pure perte.

[†] On entend par accessibilité *in vitro* la disponibilité de protocoles de test en laboratoire, de préférence quantitatifs. *In vivo*, il est utile de disposer d’un marqueur moléculaire approprié et non toxique, afin de ”visualiser” l’activité de la molécule dans l’organisme.

[‡] Il ne s’agit pas toutefois de disposer d’une connaissance *complète* de la cible, mais d’une connaissance *suffisante* afin de pouvoir mettre en place un protocole de recherche de médicaments robuste.

[§] Le screening peut être *expérimental* ou bien *virtuel.* Dans le premier cas des tests biochimiques expérimentaux mesurent l’affinité d’une molécule donnée pour la cible ; dans le second, cette affinité est estimée par des calculs numériques. Ces méthodes doivent se baser sur une référence quelconque, qui peut être la structure de la cible (si celle-ci est disponible, on parle alors de *structure-based drug design*), la structure d’un ligand de référence (*ligand-based drug design*), des jeux de données statistiques (QSAR), *etc.*

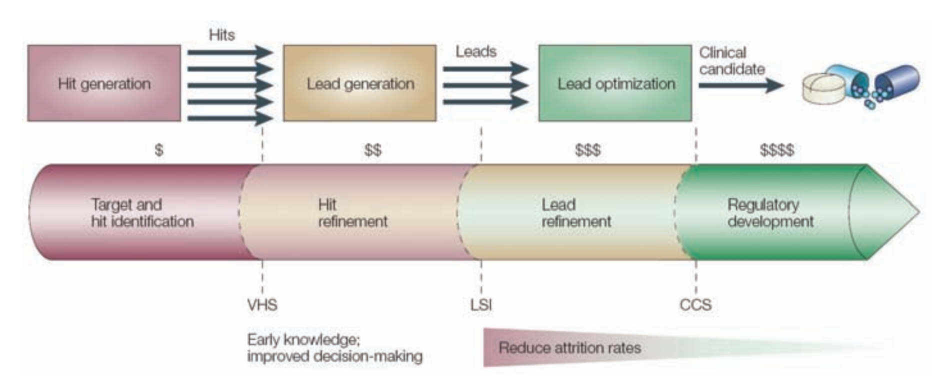
La définition que l'on fait d'un composé prometteur ou *hit* (touche) est arbitraire. Dans le cas d'un *screening*, il s'agit d'un composé qui apparaît comme interagissant significativement plus que la moyenne des composés testés sur la cible visée. Cela ne signifie en rien que ce composé sera compétitif vis-à-vis d'une référence connue, ni même qu'il aura une quelconque activité *in vivo*. Dans le cas de molécules mises au point spécifiquement, c'est la compétence du chercheur qui déterminera en premier lieu la "qualité" des *hits* mis en place.

Il est bien sûr impossible de tester par *screening* l'ensemble des molécules possibles ; celles qui le sont font partie de banques de molécules préexistantes[*], ou bien sont définies spécifiquement par rapport à la cible, *screening* "aveugle" d'une part, test de structures déterminées rationnellement d'autre part.

Divers facteurs permettent de caractériser l'utilité dans le cadre du *drug design* d'une base de molécule utilisée pour le *screening*. La diversité chimique doit être importante, et surtout les molécules doivent avoir les caractéristiques physico-chimiques caractérisant un médicament. Des règles élémentaires permettent de filtrer rapidement bon nombre de molécules ne répondant pas à ce dernier critère hors des banques (avec une probabilité faible d'exclusion de molécules valables). Les plus connues sont celles de Lipinski (*rule of five*[†]). [79] La mesure et la prédiction de la toxicité sont des problèmes plus complexes, mais qui doivent être abordés, en particulier afin d'améliorer la qualité d'une banque de molécules "générique"[‡].

Mise au point et optimisation de composés actifs spécifiques (leads)

On définit un *lead* (tête de série) comme une molécule (ou structure de base d'un ensemble de molécules) qui non seulement présente une activité significative pour la cible (*hit*), mais qui en plus est sélective pour celle-ci lors d'un test expérimental. Cette définition est arbitraire et peut varier d'un protocole de sélection à un autre. De façon générale, à partir d'un nombre important de molécules de départ, on effectue au moins trois filtrages successifs : le premier sert à identifier les *hits*, le second à sélectionner les *leads*, enfin le troisième niveau de filtrage correspond à la sélection éventuelle, après optimisations des *leads*, d'un ou plusieurs composés candidats pour les tests cliniques. Cette dernière étape correspond en général à l'emploi de techniques "manuelles" contrairement aux précédentes pour lesquelles le filtrage peut être plus ou moins automatisé.



Étapes du développement d'un médicament potentiel avant essais [80]

[*] Là encore, on ne présume en rien de la *forme* que prennent les molécules dans ces banques. Il peut s'agir d'une forme matérielle (plaques) adaptée à un screening expérimental, ou bien d'une forme immatérielle (base de données informatiques) adaptée à un screening virtuel.

[†] Les règles de Lipinski (nommées *rule of 5* à cause des paramètres et non à cause du nombre de règles) sont : (1) masse moléculaire < 500 Da, (2) donneurs de liaisons H < 5, (3) accepteurs de liaisons H (atomes N et O) < 10, (4) logP (cœfficient de partition octanol/eau) < 5 (indique que la molécule est capable de franchir les membranes biologiques).

[‡] On entend par là une banque de taille raisonnable destinée à être employée de façon routinière sur une cible quelconque. Une telle banque se doit d'avoir une diversité chimique importante, et d'exclure des composés non *drug-like*, en premier lieu les toxiques.

Essais

La ”vie” d’un médicament s’organise en plusieurs étapes, suivant un protocole rigoureux : mise au point, essais pré-cliniques, essais cliniques (phase I à III), puis suivi post-commercialisation (phase IV).

Les essais pré-cliniques ont pour objectif d’évaluer, avant l’étude chez l’homme, la sécurité du produit (toxicité, mutagenèse, cancérogenèse…), son action sur les organes cibles, ainsi que son cycle de vie dans l’organisme (absorption, propagation, élimination).

La phase I correspond à la première administration à l’homme, sur une cinquantaine de volontaires sains durant 6 à 18 mois, afin de déterminer plus précisément le dosage du produit. La phase II, effectuée en général en milieu hospitalier, sur plusieurs centaines de malades durant 2 à 3 ans, a pour but de déterminer les conditions optimales d’administration (dose et posologie) et de les relier aux concentrations passant effectivement dans l’organisme. La phase III, pouvant être étendue à plusieurs milliers de patients, mesure l’efficacité du médicament (rapport traitement de la pathologie / effets secondaires) dans les conditions d’utilisation préconisées, en concurrence avec un placebo et d’éventuels médicaments de références. Une autorisation de mise sur le marché est délivrée en cas de succès.

Bilan financier

Les experts évaluent le coût total du développement d’un nouveau médicament à l’heure actuelle (2000) à 800 millions de $. [81] À noter que ce montant correspond aux coûts totaux investis par l’industrie pharmaceutique divisés par le nombre de médicaments franchissant toutes les étapes menant à leur commercialisation ; ce chiffre inclut donc toutes les dépenses liées au développement de molécules qui finalement n’ont pas pu être commercialisées. Le laps de temps entre le démarrage du processus de recherche et la commercialisation, lorsque la recherche pharmaceutique est couronnée de succès, est d’environ 15 ans. Dans le cas du Taxol, la découverte de la molécule remonte à 1963, tandis que l’autorisation de mise sur le marché s’est faite en 1990. [82]

La découverte de médicaments suit clairement un processus en entonnoir au cours duquel, à chaque étape, de moins en moins de molécules satisfont l’ensemble des critères successifs, et dont le taux d’échec est au final particulièrement important. Si la phase de recherche de *hits* peut comprendre de plusieurs milliers à plusieurs millions de molécules, il n’est même pas certain qu’une seule de ces molécules passe la phase des essais pré-cliniques. Lorsqu’une molécule arrive à ce stade, la probabilité qu’elle doive être abandonnée suite aux essais cliniques est encore élevée. Une étude statistique mentionne 12 succès, c’est-à-dire 12 molécules dont l’effet thérapeutique supposé a été prouvé par les tests cliniques, sur 209 molécules candidates. [83]

Il est de plus crucial de noter que les coûts croissent exponentiellement d’une étape à l’autre, si bien que les progrès conceptuels des phases pré-cliniques, durant lesquelles les méthodes de simulations informatiques jouent un rôle crucial [84], doivent concentrer une bonne part des efforts des chercheurs et des industriels.[*]

Enfin, les médicaments actuellement sur le marché ne ciblent qu’une toute petite partie de l’organisme : moins de 500 biomolécules. [85] Les méthodes de recherche pharmaceutiques classiques ne sont plus efficaces, les cibles étant de plus en plus complexes et spécifiques. Dans ce contexte, l’apport des techniques de la bioinformatique, permettant une réduction des coûts et un traitement plus poussé en complément des approches expérimentales, prend naturellement un poids de plus en plus important.

[*] On peut rajouter ici que si l’industrie pharmaceutique milite fréquemment pour une simplification et surtout un racourcissement des processus de validation cliniques des médicaments, il s’agit plus pour elle d’un problème économique que d’un enjeu scientifique. L’amélioration des protocoles de sélection de candidats antérieurement aux essais cliniques est, à l’inverse, un problème méthodologique purement scientifique.

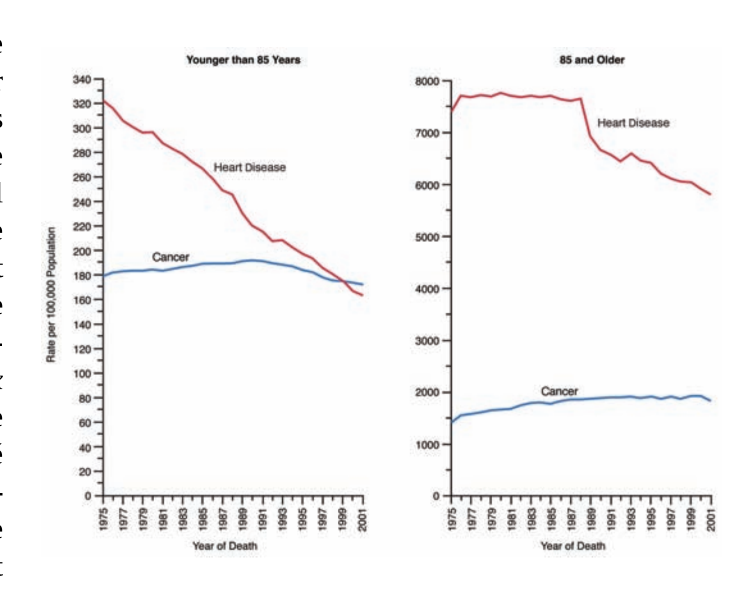
La recherche de nouveaux médicaments anti-cancer

Durant de nombreuses années, la communauté scientifique a pensé que le décryptage du génome pourrait être à l'origine d'une révolution scientifique. Si tel n'a pas été le cas (les efforts sont désormais tournés vers l'interactome), un défi scientifique de taille est posé : comment interpréter et exploiter de façon utile la masse des données obtenues ? **Le cancer étant la maladie génétique par excellence, il constitue un axe de recherche idéal pour l'exploitation des connaissances de la génomique.**

Un grand nombre d'approches désignées sous le terme d'oncologie moléculaire ont permis l'identification d'un nombre important de gènes impliqués dans les processus de prolifération cancéreuse. Le cancer étant la résultante d'un ensemble de mutations génétiques, il ne s'agit pas tant de caractériser de façon particulièrement détaillée ces anomalies, mais plutôt de détecter celles qui surviennent de façon récurrente. De nombreuses études basées à l'origine sur une telle approche et visant à la mise au point de molécules anti-cancer spécifiques sont actuellement conduites, ce qui souligne l'intérêt de la communauté scientifique. [86] Des méthodes de recherche innovantes sont susceptible d'en résulter.

La faible spécificité des traitements anti-cancer actuels[*], tels que la chimiothérapie et la radiothérapie, massivement employées, pose problème. La mise au point d'une nouvelle classe de médicaments spécifiquement anti-cancer est, dans ce contexte, particulièrement prometteuse, car porteuse à la fois d'une plus grande efficacité et d'une plus faible toxicité.

Enfin, au fur et à mesure que l'espérance de vie augmente dans les sociétés modernes, le cancer devient de plus en plus fréquent, et ses implications médicales et sociales sont, comme nous l'avons vu, un problème majeur. Il constitue aussi pour l'industrie pharmaceutique un marché émergeant particulièrement prometteur. Le cancer est déjà la première cause de mortalité dans les pays développés (*ci-contre : évolution des cause de mortalité aux Etats-Unis* [2]). On estime que d'ici à 2008, le cancer deviendra aussi le premier marché mondial pour l'industrie pharmaceutique. Celle-ci semble effectivement consentir à cet axe de recherche des moyens particulièrement importants.[†]



En résumé, **les enjeux de la recherche d'une nouvelle classe de médicaments anti-cancer sont aussi bien scientifiques que médicaux et commerciaux.**

[*] Chimiothérapie comme radiothérapie s'attaquent à la réplication de l'ADN et à la division cellulaire, et sont efficaces contre le cancer lorsqu'ils sont dirigés vers les zones de l'organisme hébergeant les tumeurs. Il s'agit de traitements "anti-prolifération" et non spécifiquement "anti-cancer", qui s'accompagnent d'effets secondaires toxiques non négligeables.

[†] On peut toutefois émettre quelques réserves sur une telle réalité. Des annonces concernant les "médicaments du futur" agissant sur des pathologies aussi complexes que les cancers avec la même facilité que des traitements antiviraux sont certainement profitables sous un angle publicitaire, et les acteurs de l'industrie pharmaceutique communiquent volontiers sur le sujet. On peut envisager que l'effort de recherche réellement consenti soit moindre, étant donné que ces entreprises dépensent désormais plus d'argent dans leurs départements marketing que dans la R&D.

Références bibliographiques

1. Pierce G.B., Shikes R. and Fink L.M. *Cancer: A problem of developmental biology*. (1978) Prentice-Hall, Englewood Cliffs, N.J.

2. Jemal A., Murray T., Ward E., Samuels A., Tiwari R.C., Ghafoor A., Feuer E.J. and Thun M.J. Cancer statistics, 2005. *CA: A Cancer Journal for Clinicians* **55** (2005) 10-30.

3. *Cancer - Pronostics à long terme. Expertise collective Inserm*. (2005) Les éditions Inserm.

4. Pinell P. How do cancer patients express their points of view? *Sociology of Health & Illness* **9**, issue 1 (1987) 25-44.

5. Moulin P. Imaginaire social et cancer. *Revue Francophone de Psycho-Oncologie* **4**, issue 4 (2005) 261-267.

6. Pardee A.B., Dubrow R., Hamlin J.L. and Kletzen R.F. Animal cell cycle. *Annual Review of Biochemistry* **47** (1978) 715-750.

7. Yanishevsky R.M. and Stein G.H. Regulation of the cell cycle in eucaryotic cells. *International Review of Cytology - A Survey of Cell Biology* **69** (1981) 223-259.

8. Sporn M.B. The war on cancer. *Lancet* **347** (1996) 1377-1381.

9. Cantley L.C., Auger K.R., Carpenter C., Duckworth B., Graziani A., Kapeller R. and Soltoff S. Oncogenes and signal transduction. *Cell* **65**, issue 5 (1991) 914.

10. Foulds L. *The experimental study of tumor progression*. I-III (1954) Academic press, London.

11. Nowell P.C. The clonal evolution of tumor cell populations. *Science* **194** (1976) 23-28.

12. Land H., Parada L.F. and Weinberg R.A. Cellular oncogens and multistep carcinogenesis. *Science* **222** (1983) 771-778.

13. Hanahan D. and Weinberg R.A. The hallmarks of cancer. *Cell* **100** (2000) 57-70.

14. Weinberg R.A. Tumor suppressor genes. *Science* **254** (1991) 1138-1146.

15. Greenblatt M.S., Bennett W.P., Hollstein M. and Harris C.C. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Research* **54**, issue 18 (1994) 4855-4878.

16. Fedi P., Tronick S.R. and Aaronson S.A. Growth factors, in *Cancer Medicine*, eds. J.F. Holland, R.C. Bast, D.L. Morton, E. Drei, D.W. Kufe, and R.R. Weichselbaum (1997) Williams and Wilkins, Baltimore, MD, p. 41-64.

17. Hunter T. Oncoprotein networks. *Cell* **88** (1997) 333-346.

18. Kerr J.F., Wyllie A.H. and Currie A.R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. **26** (1972) 239-257.

19. Kerr J.F., Winterford C.M. and Harmon B.V. Apoptosis. Its significance in cancer and cancer therapy. *Cancer* **73**, issue 8 (1994) 2013-2026.

20. Wright W.E., Pereira-Smith O.M. and Shay J.W. Reversible cellular senescence: implications for immortalization of normal human diploid fibroblasts. *Molecular and Cellular Biology* **9** (1989) 3088-3092.

21. Shay J.W. and Bacchettti S. A survey of telomerase activity in human cancer. *European Journal of Cancer* **33**, issue 5 (1997) 787-791.

22. Levine A.J. p53, the cellular gatekeeper for growth and division. *Cell* **88** (1997) 323-331.

23. Heidelberger C., Freeman A.E., Pienta R.J., Sivak A., Bertram J.S., Casto B.C., Dunkel V.C., Francis M.W., Kakunaga T., Little J.B. and Schechtman L.M. Cell transformation by chemical agents--a review and analysis of the literature. A report of the U.S. Environmental Protection Agency Gene-Tox Program. *Mutation Research* **114**, issue 3 (1983) 283-385.

24. Ashby J. and Tennant R.W. Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the US NTP. *Mutation Research* **257**, issue 3 (1991) 229-306.

25. Lynch H.T., Fusaro R.M. and Lynch J. Heredity cancer in adults. *Cancer Detection and Prevention* **19**, issue 3 (1995) 219-233.

26. Lichtenstein P., Holm N.V., Verkasalo P.K., Iliadou A., Kaprio J., Koskenvuo M., Pukkala E., Skytthe A. and Hemminki K. Environmental and heritable factors in the causation of cancer - Analyses of cohorts of twins from Sweden, Denmark and Finland. *New England Journal of Medicine* **343**, issue 2 (2000) 78-85.

27. Duesberg P.H. and Vogt P.K. Differences between the ribonucleic acids of transforming and nontransforming avian tumor viruses. *Proceedings of the National Academy of Sciences USA* **67**, issue 4 (1970) 1673-1680.

28. zur Hausen H. Papillomaviruses and cancer: From basic studies to clinical application. *Nature Reviews Cancer* **2** (2002) 342-350.

29. Påhlman L., Beger H. and Kroon B. The place of surgical oncology in general surgery. *European Journal of Surgical Oncology* **25** (1999) 619-621.

30. Dobbs J., Barrett A. and Ash D. *Practical radiotherapy planning*. 3rd ed. (1999) Arnold, London.

31. Cairns J., Boyle D. and Frei E. Cancer chemotherapy. *Science* **220**, issue 4594 (1983) 252-256.

32. Grady D., Rubin S.M., Petitti D.B., Fox C.S., Black D., Ettinger B., Ernster V.L. and Cummings S.R. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Annals of Internal Medicine* **117**, issue 12 (1992) 1016-1037.

33. Muss H.B. Endocrine therapy for advancer breast cancer: A review. *Breast Cancer Research and Treatment* **21**, issue 1 (1992) 15-26.

34. Azria D., Lemanski C., Zouhair A., Gutowski M., Belkacémi Y., Dubois J.B., Romieu G. and Ozsahin M. Hormonoradiothérapie adjuvante concomitante des cancers du sein: état de l'art. *Cancer / Radiothérapie* **8** (2004) 188-196.

35. Dachs G.U., Dougherty G.J., Stratford I.J. and Chaplin D.J. Targeting gene therapy to cancer: a review. *Oncology Research* **9** (1997) 313-325.

36. Pardoll D.M. Cancer vaccines. *Trends in Pharmacological Sciences* **14**, issue 5 (1993) 202-208.

37. Parmiani G., Castelli C., Dalerba P., Mortarini R., Rivoltini L., Marincola F.M. and Anichini A. Cancer immunotherapy with peptide-based vaccines: What have we achieved? Where are we going? *Journal of the National Cancer Institute* **94**, issue 11 (2002) 805-818.

38. Adam J.K., Odhav B. and Bhoola K.D. Immune responses in cancer. *Pharmacology & Therapeutics* **99** (2003) 113-132.

39. Berzofsky J.A., Terabe M., Oh S., Belyakov I.M., Ahlers J.D., Janik J.E. and Morris J.C. Progress on new vaccine strategies for the immunotherapy and prevention of cancer. *Journal of Clinical Investigation* **113**, issue 11 (2004) 1515-1525.

40. Rosenberg S.A., Yang J.C. and Restifo N.P. Cancer immunotherapy: moving beyond current vaccines. *Nature Medicine* **10**, issue 9 (2004) 909-915.

41. Nam N.-H. and Parang K. Current targets for anticancer drug discovery. *Current Drug Targets* **4**, issue 2 (2003) 159-179.

42. Dancey J. and Sausville E.A. Issues and progress with protein kinase inhibitors for cancer treatment. *Nature Reviews Drug Discovery* **2** (2003) 296-313.

43. Laurence V. and Espié M. Effets secondaires de la chimiothérapie. *Le Généraliste* **2118** (2001).

44. Boutin J.A. Tyrosine proteine kinase inhibition and cancer. *International Journal of Biochemistry* **26**, issue 10-11 (1994) 1203-1226.

45. Awada A., Cardoso F., Atalay G., Giuliani R., Mano M. and Piccart M.J. The pipeline of new anticancer agents for breast cancer treatment in 2003. *Critical Reviews in Oncology / Hematology* **48** (2003) 45-63.

46. Sastry L., Cao T. and King C.R. Multiple Grb2-protein complexes in human cancer cells. *International Journal of Cancer* **70** (1997) 208-213.

47. Buolamwini J.K. Novel anticancer drug discovery. *Current Opinion in Chemical Biology* **3** (1999) 500-509.

48. Clark G.J. and Der C.J. Aberrant function of the Ras signal transduction pathway in human breast cancer. *Breast Cancer Research and Treatment* **35**, issue 1 (1995) 133-144.

49. Garbay C., Liu W.-Q., Vidal M. and Roques B.P. Inhibitors of Ras signal transduction as antitumor agents. *Biochemical Pharmacology* **60** (2000) 1165-1169.

50. Sebolt-Leopold J.S. Development of anticancer drugs targeting the MAP kinase pathway. *Oncogene* **19**, issue 56 (2000) 6594-6599.

51. Boldt S., Weidle U.H. and Kolch W. The role of MAPK pathways in the action of chemotherapeutic drugs. *Carcinogenesis* **23**, issue 11 (2002) 1831-1838.

52. Kaelin Jr. W.G. Choosing anticancer drug targets in the postgenomic era. *Journal of Clinical Investigation* **104**, issue 11 (1999) 1503-1506.

53. Bange J., Zwick E. and Ullrich A. Molecular targets for breast cancer therapy and prevention. *Nature Medicine* **7**, issue 5 (2001) 548-552.

54. Watson J.D. and Crick F.H.C. Genetic implications of the structure of deoxyribonucleic acid. *Nature* **171** (1953) 964-967.

55. Watson J.D. and Crick F.H.C. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **171** (1953) 737-738.

56. Crick F.H.C. The structure of the hereditary material. *Scientific American* **191**, issue 4 (1954) 54-61.

57. Yanofsky C. Gene structure and protein structure. *Scientific American* **216**, issue 5 (1967) 80-94.

58. Crick F.H.C. On protein synthesis. *Symposia of the Society for Experimental Biology* **12** (1958) 138-163.

59. Crick F.H.C., Barnett L., Brenner S. and Watts-Tobin R.J. General nature of the genetic code for proteins. *Nature* **192** (1961) 1227-1232.

60. Brenner S., Jacob F. and Meselson M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190** (1961) 576-581.

61. Lake J.A. The ribosome. *Scientific American* **245**, issue 2 (1981) 84-97.

62. Anfinsen C.B. Principles that govern the folding of protein chains. *Science* **181**, issue 4096 (1973) 223-230.

63. Richards F.M. The protein folding problem. *Scientific American* **264**, issue 1 (1991) 54-57.

64. James L.C. and Tawfik D.S. Conformational diversity and protein evolution - a 60-year-old hypothesis revisited. *Trends in Biochemical Sciences* **28**, issue 7 (2003) 361-368.

65. Figeys D. Combining different 'omics' technologies to map and validate protein-protein interaction in humans. *Briefings in functional genomics and proteomics* **2**, issue 4 (2004) 357-365.

66. Uetz P. and Finley Jr. R.L. From protein networks to biological systems. *FEBS Letters* **579** (2005) 1821-1827.

67. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. and Bourne P.E. The Protein Data Bank. *Nucleic Acids Research* **28**, issue 1 (2000) 235-242.

68. Smith Schmidt T. Banking on structures. *BioIT World* **1**, issue 8 (2002).

69. Goto S., Okuno Y., Hattori M., Nishioka T. and Kanehisa M. LIGAND: Database of chemical compounds and reactions in biological pathways. *Nucleic Acids Research* **30** (2002) 402-404.

70. Bohacek R.S., McMartin C. and Guida W.C. The art and practice of structure-based drug design: a molecular modelling perspective. *Medicinal Research Reviews* **16** (1996) 3-50.

71. Dobson C.M. Chemical space and biology. *Nature* **432** (2004) 824-828.

72. Ellis R.J. and Minton A.P. Join the crowd. *Nature* **425** (2003) 27-28.

73. Chanda S.K. and Caldwell J.S. Fulfilling the promise: drug discovery in the post-genomic era. *Drug Discovery Today* **4** (2003) 168-174.

74. Anderson A.C. The process of structure-based drug design. *Chemistry & Biology* **10** (2003) 787-797.

75. Wlodawer A. and Vondrasek J. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annual Review of Biophysics and Biomolecular Structure* **27** (1998) 249-284.

76. Lundstrom K. Structural genomics of GPCRs. *Trends in Biotechnology* **23**, issue 2 (2005) 103-108.

77. Hénin J., Maigret B., Tarek M., Escrieut C., Fourmy D. and Chipot C. Probing a model of a GPCR/ligand complex in an explicit membrane environment: The human cholecystokinin-1 receptor. *Biophysical Journal* **90** (2006) 1232-1240.

78. Ooms F. Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Current Medicinal Chemistry* **7** (2000) 141-158.

79. Lipinski C.A., Lombardo F., Dominy B.W. and Feeney P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Discovery Reviews* **23**, issue 1 (1997) 3-25.

80. Bleicher K.H., Böhm H.-J., Müller K. and Alanine A.I. Hit and lead generation: beyond high-throughput screening. *Nature Reviews Drug Discovery* **2**, issue 5 (2003) 369-378.

81. DiMasi J.A., Hansen R.W. and Grabowski H.G. The price of innovation: new estimates of drug development costs. *Journal of Health Economics* **22** (2003) 151-185.

82. Rowinsky E.K. and Donehower R.C. Drug therapy: paclitaxel (Taxol). *New England Journal of Medicine* **332** (1995) 1004-1114.

83. Nygren P. and Larsson R. Overview of the clinical efficacy of investigational anticancer drugs. *Journal of Internal Medicine* **253** (2003) 46-75.

84. Jorgensen W.L. The many roles of computation in drug discovery. *Science* **303**, issue 5665 (2004) 1813-1818.

85. Drews J. Drug discovery: A historical perspective. *Science* **287** (2000) 1960-1964.

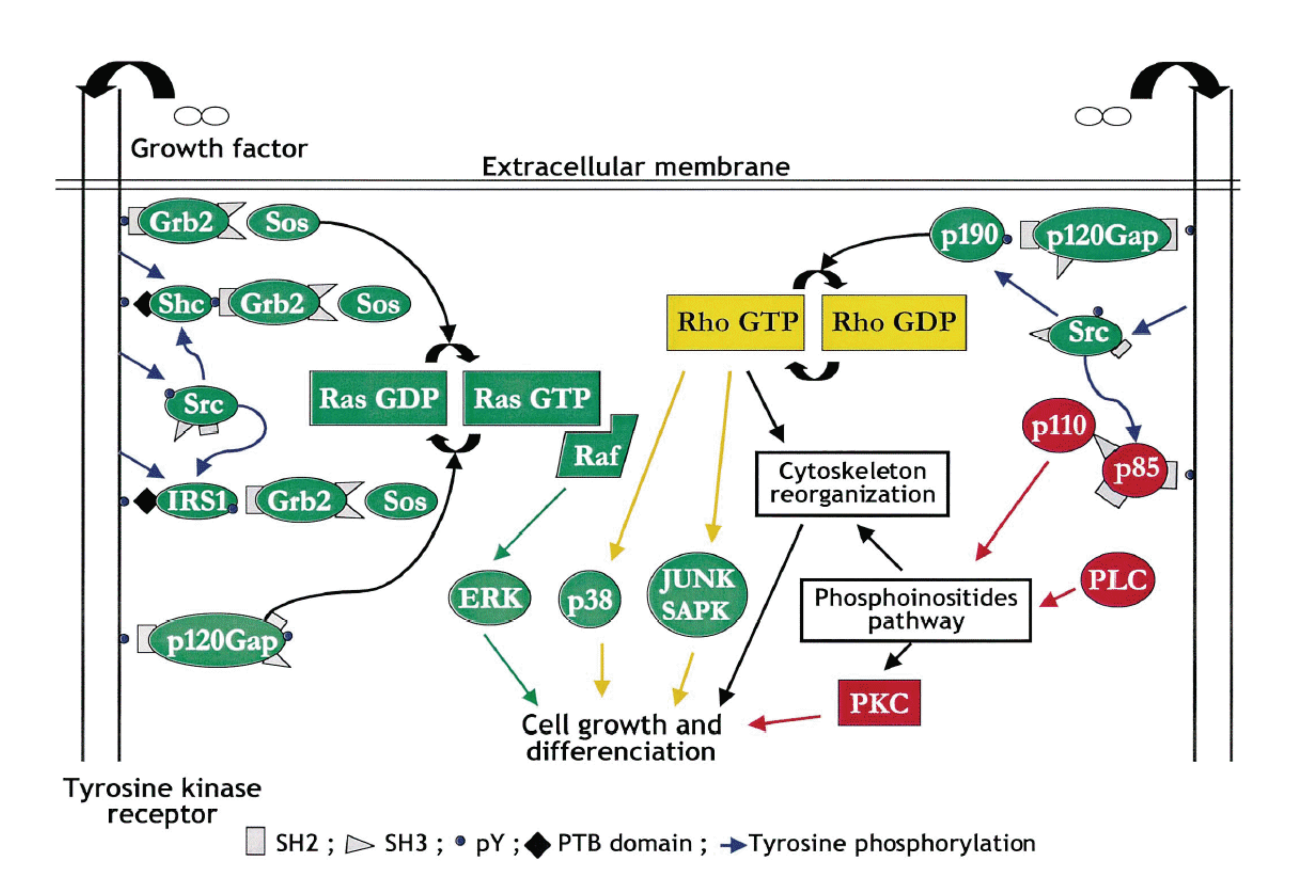
86. Garrett M.D. and Workman P. Discovering novel chemotherapeutic drugs for the third millenium. *European Journal of Cancer* **35**, issue 14 (1999) 2010-2030.

Cible

Description

La voie de signalisation Ras-MAPK

Dans un organisme vivant, une voie de signalisation correspond à une cascade d'interactions entre protéines dont la finalité est la propagation d'un signal biologique. De tels processus ont lieu à l'échelle de la milliseconde, voire de la seconde. Ils sont localisés aussi bien à l'extérieur qu'à l'intérieur des cellules ; ceux qui sont à la fois extra- et intra-cellulaires permettent aux cellules de communiquer avec leur environnement. Souvent, un stimulus mineur au début de la voie de signalisation donne lieu à une réponse biologique bien plus importante. L'ensemble des voies de signalisation constitue un circuit logique définissant en grande partie le fonctionnement d'un organisme. L'identification, la caractérisation et la classification fonctionnelle des différentes voies de signalisation d'un organisme constitue donc un aspect crucial dans l'étude de son interactome.



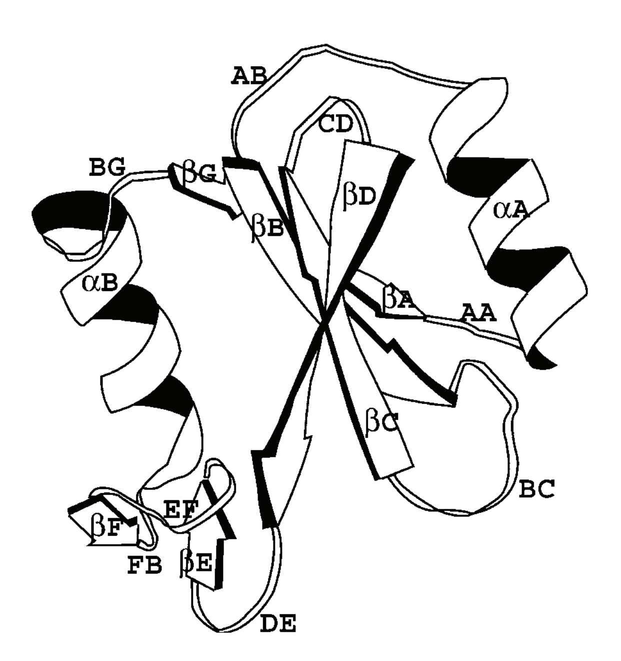
Fonctionnement simplifié des mécanismes de croissance et différenciation cellulaire incluant la voie de signalisation Ras-MAPK (partie gauche) [1]

La voie de signalisation Ras-MAPK[*] [2-6] correspond à une cascade d'interactions intra-cellulaires essentielle dans la croissance, la régulation et la différentiation des cellules [7, 8]. Elle semble également être impliquée dans des processus biologiques variés tels que le fonctionnement de la mémoire à long terme. [9] Elle suscite un grand intérêt scientifique, car les mutations du gène Ras[†] se signalent parmi les mutations les plus fréquemment observées dans les cas de cancers humains [6, 10, 11]. De plus, certaines de ces mutations sont en particulier reliées à l'activité de Ras-MAPK, ce qui fait d'elle une cible thérapeutique [12, 13]. Au sein de Ras-MAPK, on trouve deux formes complexées de Ras : Ras GDP et Ras GTP. De façon générale, la protéine Ras a besoin d'être modifiée, souvent par complexation, pour acquérir une activité biologique. Dans Ras-MAPK, seule la forme Ras GTP est active : le passage de Ras GDP à Ras GTP provoque un changement conformationnel qui expose un site actif de Ras [14] pour Raf [15, 16], ce qui permet de poursuivre la cascade d'interaction de Ras-MAPK. Cette transformation est favorisée, entre autres, par l'intervention de complexes de la protéine-adaptatrice Grb2. [17, 18]

Les domaines SH2

On désigne par *domaine* une famille de séquences d'acides aminés présente sur un nombre important de protéines[‡] et dont les membres présentent un très fort taux d'homologie. Les domaines correspondent à des codes génétiques généralement fortement conservés au cours de l'évolution.

Les domaines SH2 (Src homology 2) [19] sont des séquences d'environ 100 acides aminés d'abord observées en commun dans les oncoprotéines Src et Fps [20], puis ensuite dans une grande variété de protéines intra-cellulaires [21], telles que Fyn, PI3K/p85α, p56-Lck, Syk, Sap, Grb2... soit plus de 100 protéines dans le génome humain. [22] La conformation des domaines SH2 correspond à deux hélices α encadrant plusieurs feuillets ß centraux anti-parallèles (*voir ci-contre*). La séquence de la plupart[§] des domaines SH2 suit une géométrie ßαßßßßαß ; on décompose alors la séquence en zones géométriques notées successivement ßA, AA', αA, AB, ßB, BC, ßC, CD, ßD... [23]



[*] Ras est un oncogene qui a été historiquement isolé sur des rats et est associé au virus Sarcoma, d'où son sigle. MAPK signifie "mitogen-activated protein kinase". La voie de signalisation Ras-MAPK est parfois désignée différemment : Ras kinase, Ras-Erk, Ras-Raf-Mek-Erk-MAPK…

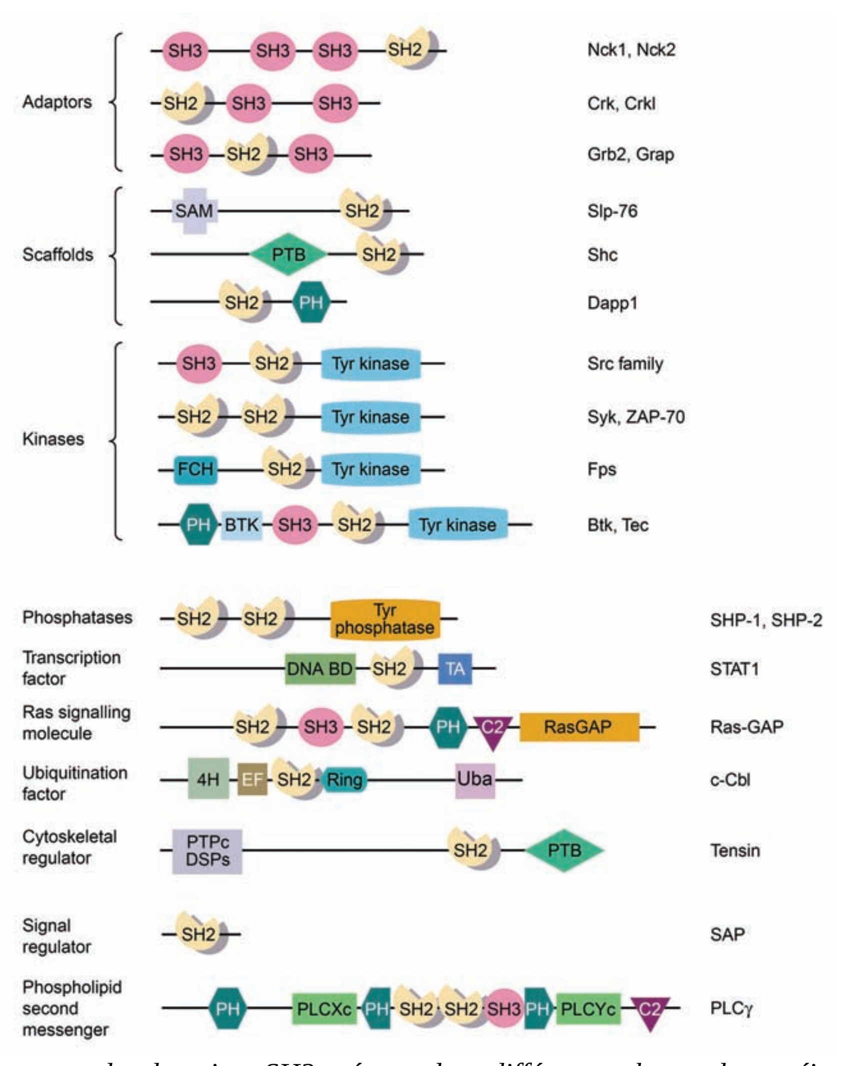
[†] Ras peut désigner aussi bien la protéine qui intervient dans la voie de signalisation Ras-MAPK que le gène qui code pour la production de cette protéine. Ici les mutations du gène dont on parle sont bien entendu corrélées à des mutations sur la protéine (concernant principalement les résidus 12, 13 et 61).

[‡] Ces protéines n'ont pas nécessairement les mêmes fonctions, et le domaine non plus, s'il est exposé.

[§] La séquence des domaines SH2 de Cbl et STAT1 a une géométrie αßßßα.

La particularité des domaines SH2 réside dans une très haute affinité pour des petites séquences protéiques débutant par un résidu tyrosine phosphorylé (pTyr). [24, 25] Chaque famille de domaines SH2 reconnaît préférentiellement des séquences plus spécifiques, selon la nature des résidus à proximité immédiate de la phosphotyrosine.* [24] Les domaines SH2, en général, se lient aux résidus pTyr présents sur les récepteurs tyrosine kinase [26] et régulent ainsi leur activité [27], ce qui les relie en particulier à différentes cascades de réactions (dont Ras-MAPK) contrôlant la division cellulaire. [28, 29]

Les domaines SH2 sont ainsi des cibles intéressantes pour la mise au point de régulateurs de la division cellulaire, classe de molécules potentiellement anticancéreuses. De plus, la surexpression de protéines contenant des domaines SH2 (impliquant une dérégulation des voies de signalisation correspondantes) caractérise souvent le développement de tumeurs cancéreuses, ce qui fait des domaines SH2, tout comme Ras-MAPK, des cibles thérapeutiques de choix dans la recherche pharmaceutique contre le cancer. [1, 12, 13, 30]



Positionnement des domaines SH2 présents dans différentes classes de protéines. [31]
Les fonctions correspondantes dans l'organisme sont indiquées à gauche.

* Par exemple, Src SH2 reconnaît préférentiellement la séquence pYEEI, PI3K/p85α pY(M/V)XM (X = acide aminé quelconque), Crk pYXXP, Grb2 pYXN, Cbl D(N/D)XpYXXX(P/F), STAT pYXXQ, SAP (T/S)IpYXX(I/V), Syk pYEXL, Sh-PTP1 (I/L/V/S)XpYXX(I/L/V), Sh-PTP2 (V/S)XpY(I/V)X(I/L/V)X(W/F), Vav pYMEP, *etc.*

La protéine Grb2



Description et fonction

Grb2 (growth factor receptor bound protein 2) est une petite protéine globulaire d'environ 200 résidus, présente en milieu intra-cellulaire, et simplement constituée d'un domaine SH2 compris entre deux domaines SH3. [32-35]

Les fonctions que peuvent prendre les protéines sont déterminées par leur réactivité chimique, qui dépend en grande partie de leur géométrie, et est définie par les séquences particulières d'acides aminés. La diversité fonctionnelle des protéines est particulièrement importante.[*] Grb2 est une protéine adaptateur : elle permet de mettre en relation des protéines, généralement de taille plus importante. Grb2 possède une activité significative dans un nombre varié de processus biologiques [36-39]. En particulier, lorsqu'elle est complexée avec Sos par ses domaines SH3[†] [40], Grb2 se lie très efficacement par son domaine SH2 avec des récepteurs tyrosine kinase phosphorylés tels que le facteur de croissance EGF-R [41], Shc [42, 43] ou FAK [44]. Le complexe Grb2/Sos active alors la voie de signalisation Ras-MAPK [18] dont nous avons décrit précédemment les caractéristiques principales, en se positionnant comme intermédiaire entre l'expression initiale de facteurs de croissance et l'activation de Ras. Il est important de noter que l'inhibition de l'activité de Grb2, via son domaine SH2 ou ses deux domaines SH3, constituerait un blocage du mécanisme de la voie de signalisation Ras-MAPK, et plus généralement de différents processus de régulation cellulaires. [45]

Intérêt pharmaceutique

Le domaine SH2 de Grb2 se situe dans un contexte multiple particulièrement intéressant du point de vue pharmacologique : les mécanismes de croissance cellulaire en général et la voie de signalisation Ras-MAPK en particulier, les protéines à activité tyrosine kinase, ainsi que les domaines SH2, constituent tous, à différents niveaux, des systèmes activement étudiés dans l'optique de la mise au point de médicaments innovants.

Un grand nombre de voies de signalisation relatives à la croissance cellulaire sont régulées par des réactions de phosphorylation (grâce aux kinases), de déphosphorylation (via les phosphatases), souvent grâce à un équilibre entre ces deux processus antagonistes. [46] En effet, le changement de l'état de phosphorylation de certaines protéines au niveau de résidus tyrosine, sérine et thréonine principalement constitue alors un "interrupteur" pour le signal propagé par la voie de signalisation. Celle-ci possède dans de nombreux cas une étape qui implique la reconnaissance moléculaire d'un résidu pTyr, tâche pour laquelle sont spécialisés les domaines SH2. [26]

---

[*] Certaines protéines extra-cellulaires ont surtout un rôle mécanique, anatomique. Ainsi, la collagène, dont la structure est une triple hélice extrêmement stable, forme par empilement les tendons, fibres musculaires particulièrement résistantes à la tension. L'élastine, à l'inverse, est une constituante essentielle des tissus, fournissant l'élasticité indispensable au fonctionnement artériel. De nombreuses protéines sont des *enzymes* : leur rôle est de catalyser des réactions biochimiques. Des protéines peuvent également avoir une spécialisation biologique, par exemple en constituant le système immunitaire (anticorps) ou bien en transcrivant l'information génétique (ribosome). On qualifie de *transporteurs* les protéines dont le rôle est la diffusion dans l'organisme de substrats, telle l'hémoglobine qui transporte l'oxygène dans le sang. De nombreuses autres catégories fonctionnelles de protéines peuvent être définies.

[†] Le domaine SH3 de Grb2 reconnaît spécifiquement des motifs peptidiques riches en proline, tels que ceux présents sur Sos.

Une grande variété de pathologies peuvent être corrélées au niveau de l'interactome à la dérégulation d'une voie de signalisation bien précise. En particulier parmi celles qui ont trait à la régulation de la croissance cellulaire, la surexpression des signaux protéine kinase est particulièrement fréquente. Leur inhibition constitue ainsi une cible thérapeutique extrêmement intéressante [47-49] à laquelle l'industrie pharmaceutique accorde une grande importance (*ci-contre : sélection d'inhibiteurs de protéines kinase en phase de test clinique* [50]). Au niveau moléculaire, ces recherches visent en premier lieu à identifier les protéines de la voie de signalisation visée les plus impliquées dans l'expression exagérée du signal biologique. Un nombre important de liens de causalité entre différentes pathologies et une protéine contenant un domaine SH2 a ainsi été établi. [51] L'inhibition de la fonction de ce dernier (la reconnaissance des séquences présentant un résidu pTyr) est alors susceptible de constituer un traitement médical efficace et spécifique. [1, 19, 52]. La détermination d'inhibiteurs pour les différentes familles de domaines SH2 constitue donc un sujet de recherche d'intérêt pharmaceutique particulièrement suivi. [19, 53]

Kinase Target	Agent	Trial (Disease)	Sponsor
Tyrosine kinases			
ABL (c-Kit, PDGFR)	Gleevec (STI-571)	Approved (CML)	Novartis
EGFR	ZD1839 (Iressa)	Approved (lung cancer)	AstraZeneca
	OSI-774	Phase III (cancer)	OSI/Roche/Genentech
	IMC-C225 (mAb)	Phase III (cancer)	ImClone
	ABX-EGF (mAb)	Phase II (cancer)	Abgenix
	MDX-447 (mAb)	Phase I (cancer)	Merck KgaA
	EMD 72000 (mAb)	Phase I (cancer)	Merck KgaA
	Genistein	Phase II (cancer)	NCI
	RH3 (mAb)	Phase II (cancer)	York Medical Bioscience Inc
EGFR, ERB2R	CI1033	Phase II (cancer)	Pfizer
	EKB569	Phase I (cancer)	Wyeth-Ayerst
	GW2016	Phase I (cancer)	GlaxoSmithKline
	PKI166	Phase I (cancer)	Novartis
VEGFR (PDGFR, FGFR)	SU6668	Phase I (cancer)	Pharmacia Corp
PDGFR (Flt-3)	CT53518	Phase I (cancer)	Millennium Pharmaceuticals
VEGFR	SU5416	Phase III (cancer)	Pharmacia Corp
	PTK787/ZK222584	Phase II (cancer)	Novartis/Schering-Plough
VEGFR (EGFR)	ZD6474	Phase II (cancer)	AstraZeneca
VEGFR (PDGFR)	SU011248	Phase II (cancer)	Sugen
NGFR, Trk	CEP-2583	Phase II (cancer)	Cephalon
HER-2/neu	17-AAG	Phase I (cancer)	Kosan
	Trastuzumab (mAb)	Approved (cancer)	Genetech
	2C4 (mAb)	Phase I (cancer)	Genetech
	CP-724,714	Phase I (cancer)	OSI Pharmaceuticals/Pfizer
	MDX-210 (mAb)	Phase I (cancer)	Novartis
Serine/threonine kinases			
PKC, c-Kit, PDGFR	PKC412	Phase II (cancer, retinopathy)	Novartis
PKC	ISIS 3521	Phase III (cancer)	ISIS Pharmaceuticals
	CGP41251	Phase II (cancer)	Novartis
	UCN-01	Phase I/II (cancer)	Kyowa Hakko Kogyo
	Bryostatin-1	Phase I/II (cancer)	Biotek
PKC-β	Ly333531	Phase I (cancer)	Eli Lilly
		Phase II/III (diabetic neuropathy)	
CDKs	Flavopiridol	Phase II (cancer)	Aventis
	E7070	Phase I (cancer)	EISAI
	BMS-387032	Phase I (cancer)	Bristol-Myers Squibb
	CYC202	Phase I (cancer)	Cyclacel
MEK1/2	PD184352	Phase II (cancer)	Pfizer
	U-0126	Phase I (cancer)	Promega
MLK	CEP-1347	Phase II (neurodegeneration)	Cephalon
RAF	BAY43–9006	Phase II (cancer)	Onyx Pharmaceuticals/Bayer
	ISIS5132	Phase II (cancer)	Isis pharmaceuticals
	L-779,450	Phase II (cancer)	Merck
Ras	ISIS2503	Phase II (cancer)	Isis pharmaceuticals
	SCH66336	Phase II (cancer)	Schering-Plough
	BMS214662	Phase I (cancer)	Bristol-Myers Squibb
	R115777	Phase I/II (cancer)	Johnson & Johnson
mTOR	CCI779	Phase II (cancer)	Wyeth-Ayerst
	RAD001	Phase I (cancer)	Novartis
		Phase II/II (immunosuppressant)	
	Rapamycin	Approved (immunosuppressant)	Wyeth-Ayerst
p38-MAPK	VX702	Phase II (inflammation; ACS)	Vertex Pharmaceuticals
	BIRB796	Phase III (inflammation; RA; Crohn's)	Boehringer Ingelheim
	SCIO-323	Phase I (RA; stroke; diabetes)	Scios, Inc
	SCIO-469	Phase II (RA; Crohn's)	Scios, Inc
PDK1	UCN-01	Phase I/II (cancer)	Kyowa Hakko Kogyo
JNK1–3	CC401	Phase I	Celgene

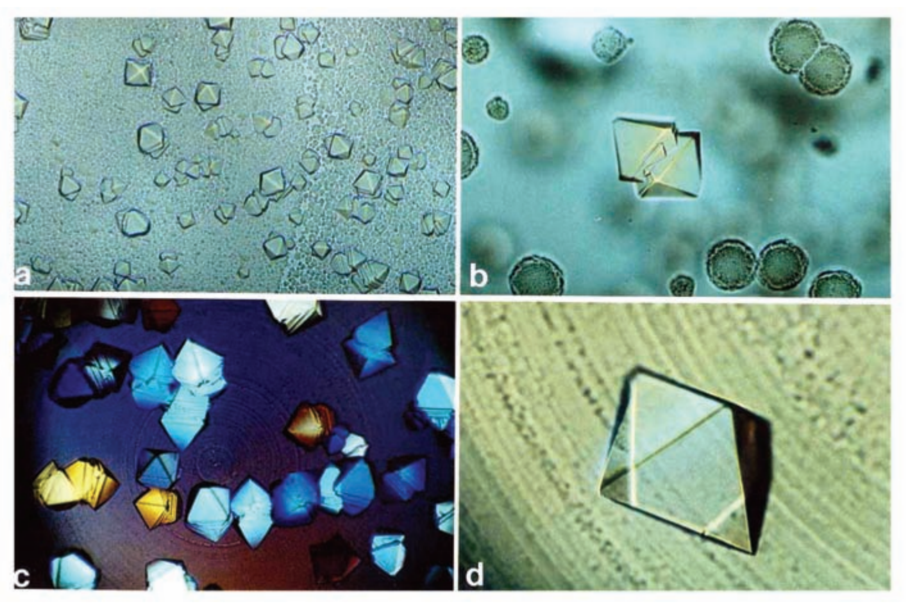
VEGFR indicates vascular endothelial growth factor receptor; PDGFR, PDGF receptor; FGFR, fibroblast growth factor receptor; CML, chronic myelogenous leukemia; RA, rheumatoid arthritis; and ACS, acute coronary syndromes. Inhibitors are of two types, monoclonal antibodies (mAbs), which are directed at the extracellular domain of various receptor tyrosine kinases, and small-molecule inhibitors.

Dans environ 25% des cas de cancers humains, des protéines exprimant le signal Ras sont présentes sous des formes altérées dans lesquelles elles ne sont plus régulées normalement, permettant ainsi aux cellules cancéreuses d'acquérir leur autonomie fonctionnelle. [54] La recherche d'inhibiteurs de Ras-MAPK suscite pour cette raison un très grand intérêt*. [6] Dans ce contexte, nous avons étudié plus particulièrement l'inhibition du domaine SH2 de Grb2. Il a été démontré que la prolifération anarchique des cellules caractérisant certaines formes de leucémie [55], ainsi que les cancers du sein, des ovaires, du poumon et de la prostate [56-58], est reliée à une surexpression de HER2/ErbB2, un analogue du complexe Grb2/EGF. [59] Des inhibiteurs de Grb2 SH2 sont également susceptibles de réduire, non seulement la capacité de multiplication, mais aussi la mobilité et les capacités de propagation de cellules cancéreuses. [45, 60, 61] Leur mise au point en tant qu'éventuels médicaments anti-cancer est donc pleinement justifiée. [1, 12, 13, 62-64]

* Actuellement, certains produits de la classe des inhibiteurs de farnesyltransferase, actifs sur Ras-MAPK et pouvant être prescrits par voie orale, sont testés cliniquement en tant que médicaments anti-cancer spécifiques.

Accessibilité expérimentale

Grb2, tout d'abord, est une biomolécule accessible pour l'expérimentateur ; les protocoles de cristallisation sont bien définis. [65] Des structures pour Grb2 SH2 libre [35, 66-68] et complexé avec différents ligands [68-74] sont disponibles dans la base PDB [75, 76], aussi bien par RX [35, 68-71, 73, 74] que par RMN [66, 67, 70, 72]. Cela constitue un argument de poids en faveur de l'utilisation de Grb2 comme cible thérapeutique à travers des études de modélisation moléculaire. En effet, de telles études reposent sur le modèle de départ employé ; dans le cas de Grb2, un tel modèle sera dérivé directement de structures expérimentales, dont on peut assurer la fiabilité.*



Cristaux de Grb2 [65]

L'accessibilité expérimentale à Grb2 SH2, déjà garantie sur un plan purement structural, est également assurée au niveau de la grandeur physique qui quantifie l'intérêt ou non d'un inhibiteur donné : l'affinité. En effet, un protocole standardisé, ELISA [77] est disponible et permet de mesurer *in vitro* de façon fiable et relativement simple le pouvoir d'inhibition d'un produit donné sur le fonctionnement d'un domaine SH2 quelconque dans l'organisme. Au moment de débuter cette thèse, des mesures obtenues par ce protocole pour un nombre significatif d'inhibiteurs de Grb2 étaient déjà publiées. [69, 78-83] L'étude théorique de l'inhibition Grb2 SH2 inclut la mise en place d'un protocole permettant d'estimer l'affinité d'un ligand donné ; toutefois, la valeur d'un tel protocole serait très limitée s'il n'était pas possible de valider expérimentalement ses résultats. La présence en parallèle d'un protocole expérimental robuste fournissant directement des valeurs d'affinité correspond à ce niveau à la meilleure des situations. Ajoutons enfin que parmi les équipes de chercheurs impliquées dans l'ACI ayant permis la conduite de cette thèse sont présentes différentes équipes d'expérimentateurs aptes à procéder aussi bien à la synthèse chimiques d'inhibiteurs potentiels qu'à la mesure de leur affinité par le protocole ELISA. C'est un point extrêmement positif aussi bien dans ce qu'il implique de collaboration entre chercheurs issus d'horizons scientifiques différents qu'au niveau de la variété des approches qui pourront être entreprises pour améliorer la compréhension de la cible d'étude.

Différentes techniques expérimentales autres que le protocole ELISA ont également été proposées afin de caractériser l'inhibition de la cible Grb2 SH2 [84-86]. Bien que leur emploi ne semble pas préférable si l'on cherche à obtenir des valeurs d'affinité quantitatives, elles peuvent se révéler utiles dans un contexte purement qualitatif, par exemple pour l'identification d'inhibiteurs par screening.

* Un tel cas de figure est évidemment bien plus favorable que celui qui requiert la construction par homologie d'un modèle de structure dont la qualité conditionnera celle de toutes les simulations ultérieures qui l'emploieront.

Accessibilité théorique

Nous venons de voir que Grb2 SH2 s'avérait accessible sur le plan expérimental, de par la disponibilité des structures expérimentales nécessaires à l'emploi de méthodes de la modélisation moléculaire, ainsi qu'au niveau des expériences qui pourront être à la fois conduites de façon complémentaire aux simulations et afin de valider celles-ci. Qu'en est-il de l'accessibilité de Grb2 SH2 au niveau des simulations elles-mêmes ?

Tout d'abord, d'une façon générale, nous avons à modéliser des complexes biologiques constitués d'une protéine liée à un inhibiteur. Un système moléculaire d'une telle taille ne peut pas être étudié par toutes les méthodes de la chimie théorique ; il peut toutefois être traité par un certain nombre d'entre elles. Dans le cadre de cette étude, nous avons principalement effectué des calculs par dynamique moléculaire et docking flexible. On peut remarquer que relativement à la plupart des complexes moléculaires représentatifs d'une activité biologique*, Grb2 SH2 correspond à une structure de taille modeste. Les ligands peptidiques de Grb2 SH2 sont ainsi des séquences constituées de 4 acides aminés actifs [87], ce qui est raisonnable et permet d'envisager que la diversité chimique des inhibiteurs possibles soit accessible. À moyens de calcul égaux et pour un protocole de simulation donné, il est donc à priori possible d'obtenir des résultats plus précis, ou en plus grand nombre, en comparaison avec la moyenne des cibles biologiques potentielles.

Antérieurement à cette thèse, des travaux de recherche théoriques sur Grb2 SH2 ont révélé de nombreuses informations intéressantes [70, 78, 79, 88, 89], et constituent une base de travail appréciable. Pour autant, si on dispose d'une connaissance initiale satisfaisante du domaine SH2 de Grb2, celle-ci n'apparaît pas pour autant suffisante pour résoudre toutes les interrogations s'y rapportant. On peut souligner que c'est le cas des domaines SH2 en général. [90] La cible Grb2 SH2 est toujours en cours d'investigation, traduisant son intérêt de la part de la communauté scientifique.

Notons enfin que les complexes formés par Grb2 SH2 avec des inhibiteurs présentent certains autres avantages facilitant leur modélisation. Tout d'abord, il a été démontré que le domaine SH2 de Grb2 a une activité biologique indépendante de ses deux domaines SH3 [91] ; en conséquence, ces derniers n'ont pas à être inclus dans les modèles. Ensuite, la liaison Grb2 SH2 / ligand fait intervenir une unité Grb2 SH2 et une unité du ligand ; à l'inverse, la liaison Grb2 SH3 / ligand implique en général les deux domaines SH3, ce qui complique la modélisation (au vu d'une publication récente [74], il pourrait toutefois être nécessaire de reconsidérer ce point†). Enfin, comme allons le voir, la forme et le mode de liaison des ligands de Grb2 SH2 est spécifique et bien déterminée, ce qui devrait simplifier les démarches de mises au point de nouveaux inhibiteurs.

* On peut explorer certaines structures issues de la base de données PDB pour se faire une idée. Le site http://www.rcsb.org/ présente chaque mois une section consacrée aux structures les plus remarquables – cliquer sur le lien "Molecule of the month" de la page d'acceuil.

† Celle-ci indique que certains ligands synthétiques pourraient se lier avec deux unités de Grb2 SH2. Reste à déterminer si une telle structure, observée par crystallographie RX, caractérise bien le complexe dans les conditions physiologiques. Pour ce faire, la publication de la structure RMN de ce complexe serait idéale.

État actuel des connaissances sur l'inhibition de Grb2 SH2

Ligands peptidiques

Détermination de la séquence de référence

Tous les domaines SH2 sont des récepteurs naturels pour les séquences peptidiques de type pYXXX, où pY désigne une tyrosine phosphorylée et X n'importe quel acide aminé. [92] En ce qui concerne Grb2 SH2, les séquences naturelles connues sont principalement pYINQ (ligand : EGF-R) [41, 93] et pYVNV (Shc) [42].

La première étape dans la mise au point d'inhibiteurs pour Grb2 SH2 consista à identifier la séquence peptidique optimale. Ce problème fut résolu par Songyang *et al.* qui effectuèrent un screening par chromatographie des séquences pYXXX sur un grand nombre de domaines SH2. [24, 87] Il s'avéra que les trois résidus préférentiels consécutifs à la phosphotyrosine sont spécifiques à chaque domaine SH2, bien que des similitudes permettent d'effectuer une classification des domaines SH2 en familles. De façon intéressante, il est constaté que certains domaines SH2 (dont Grb2) sont hautement sélectifs sur la position pY+2 ; d'autres (par exemple, Src) le sont pour la position pY+3. Des travaux ultérieurs montrèrent que certains domaines SH2 sont sélectifs bien en-dehors du motif pYXXX. [19] En ce qui concerne le domaine SH2 de Grb2, les séquences de type pYXNX sont préférées.

Différents screenings effectués plus spécifiquement sur Grb2 SH2 [84, 93-96] soulignèrent ensuite l'importance d'un résidu asparagine en position pY+2. De façon plus précise, des mesures calorimétriques ont démontré que sa substitution par une alanine divise l'affinité de la séquence pYVNV 2000 fois [97]. Deux études [93, 98] suggèrent toutefois que des séquences ne disposant pas d'un résidu asparagine en position pY+2 – en particulier pYQQD – sont favorables à un degré moindre, ce qui semble indiquer que la présence de l'asparagine constitue plus une contrainte conformationnelle très favorable qu'une obligation structurale. Les résultats divergent quant à la préférence sur les sites pY+1 et pY+3 ; ceux-ci ne doivent donc pas être cruciaux en regard de l'interaction sur le récepteur de Grb2 SH2, bien qu'il ait été proposé de placer un résidu chargé positivement en pY+1 afin d'augmenter la spécificité du ligand sans pour autant handicaper trop fortement son affinité. [99] À la lumière de ces informations, il est légitime de considérer la séquence naturelle pYVNV comme une référence en vue de mettre au point des inhibiteurs - peptidiques ou non - plus performants ciblant le domaine SH2 de Grb2.

Mode de liaison

Comme avec l'ensemble des domaines SH2, le récepteur du domaine SH2 de Grb2 possède deux cavités bien distinctes : la première reconnaît spécifiquement un résidu pTyr, tandis que la seconde, plus spécifique, est hydrophobe. La zone de reconnaissance pTyr est caractérisée principalement par la présence de deux arginines, R_{67} (Arg αA) et R_{86} (Arg ßB)[*] ; elle possède donc une charge +2, ce qui permet d'interpréter la force de la liaison pTyr / SH2. Le reste de la cavité est composé de trois sérines, S_{88} (Ser ßB), S_{90} (Ser BC) et S_{96} (Ser ßC). La cavité hydrophobe, quand à elle, est délimitée par K_{109} (Lys ßD) et constituée de L_{120} (Leu

[*] La numérotation correspond à l'ordre des résidus sur Grb2, tandis que la notation entre parenthèses distingue les résidus suivant la zone géométrique de SH2 à laquelle ils appartiennent (dans l'ordre, ßA, AA', αA, AB, ßB, BC…).

ßE) et W_{121} (Trp EF) qui interagissent plus faiblement avec l'asparagine de la séquence de référence pYXNX. Ce modèle est conforme à la description classique et simpliste [100] des complexes de domaines SH2, qui comprennent pour la plupart un ligand à deux "branches" liées dans deux cavités distinctes, l'une pour pTyr, et l'autre hydrophobe. La séquence du domaine SH2 de Grb2 contient également le motif "FLIRES" analogue au motif "FLVRES" souvent cité comme étant d'une importance capitale pour la reconnaissance de la phosphotyrosine* et présent sur la plupart des domaines SH2.

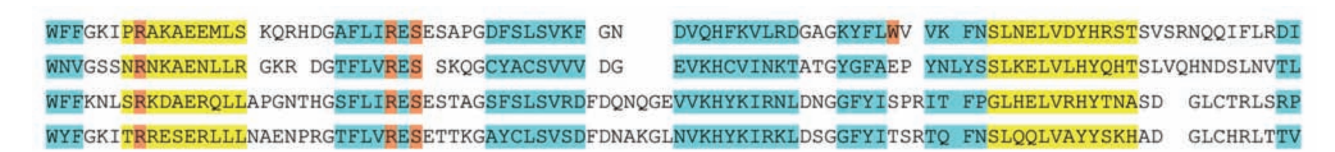
Spécificité

Grb2 SH2, comme nous l'avons vu, est doublement sélectif sur les séquences peptidiques auquel il se lie : si la présence d'un résidu pTyr est attendue car systématique avec les domaines SH2, la sélectivité pour un résidu asparagine en position pY+2 distingue Grb2 des autres protéines contenant un domaine SH2. L'analyse des structures RX de complexes peptidiques de Grb2 SH2 [68-70] comparées à celles de complexes d'autres domaines SH2 [23, 77, 101-105] permet de vérifier que cette spécificité se retrouve dans le mode de liaison correspondant.

L'analyse des structures RX montre en effet que la surface accessible du récepteur de Grb2 SH2 diffère totalement de celle des autres récepteurs SH2 par le résidu W_{121} (Trp EF), dont l'importance avait déjà été mise en évidence par mutagenèse la détermination des structures géométriques [25].

W_{121} (Trp EF), de par son encombrement, impose une contrainte géométrique spécifique à Grb2 en plus de la nécessité de la présence d'un résidu pTyr, commune à tous les domaines SH2. Seules les chaînes peptidiques présentent un tour ß en position pY+1, d'où une conformation "coudée" que seul un résidu asparagine en position pY+2 permet de garantir, ne sont pas défavorisés. [106-108]

Ainsi, une séquence ne présentant pas le motif pYXNX sera spécifiquement inactive avec Grb2 SH2.† Ainsi, la séquence pYINQ a la même activité avec Grb2 SH2 et Lck SH2, tandis que la séquence pYEEI n'est active qu'avec Lck SH2, et de façon plus poussée que pour pYINQ. [107] L'interaction d'un ligand avec un domaine SH2 peut donc se faire suivant au moins deux modes de liaison, l'un "linéaire" et l'autre "courbé", le premier étant à priori plus favorable, sauf dans le cas de Grb2 où seul le second est possible.



Séquences des domaines SH2 de (de haut en bas) Grb2, PI3K/p85α, P56-Lck et C-Src.
En jaune, résidus formant une hélice α ; en bleu, résidus formant un feuillet ß.
En orange, résidus fondamentaux pour la reconnaissance des ligands.

* L'importance de ce motif nous semble toutefois sur-estimée, car il diverge sur un certains nombre de domaines SH2, par exemple avec Cbl on a "YIFRLS", avec Syk "FLIRAR", avec SAP "YLLRDS"… sans pour autant que la sélectivité de ces domaines SH2 pour les phosphotyrosines soit remise en question. L'analyse des structures RX montre que seul le résidu arginine de "FLIRES", conservé dans tous les domaines SH2, est vraiment crucial, ainsi dans une moindre mesure la sérine et une autre arginine sur l'hélice αA, conservées dans la plupart des cas. Il est donc sans doute plus juste de définir plus généralement la spécificité-pTyr des domaines SH2 comme la résultante de la présence de plusieurs résidus arginine et sérine accessibles dans une cavité formée à la jonction de l'hélice αA et du feuillet ßB.

† On peut souvent lire que pYXNX est un motif spécifique à Grb2 SH2, ce qui n'est pas tout à fait exact. En effet, ce motif peut être actif avec n'importe quel domaine SH2 ; toutefois il l'est alors moins que d'autres motifs. Concernant Grb2, ces autres séquences sont quasiment toutes inactives. C'est bien cette inactivité qui est spécifique, et non l'activité de pYXNX.

Ligands pseudo-peptidiques ou non peptidiques optimisés

Substitution de groupes peptidiques (Novartis)

La première approche visant à découvrir des inhibiteurs plus efficaces de Grb2 SH2 consista pour une équipe de Novartis, après avoir résolu la structure RX d'un complexe de référence Grb2 SH2 / KPFpYVNV [106] (code PDB : 1ZFP), à se baser sur la petite séquence peptidique Ac-pYIN-NH_2, qui conserve une affinité micromolaire. Les optimisations de cette séquence ont ensuite consisté à tester l'effet de substituants non peptidiques [89, 109] aux différents groupes Ac_{-1} [69, 110], $pTyr_0$ [111, 112], Ile_{+1} [78, 88, 113, 114], Asn_{+2} [71, 113] et $(NH_2)_{+3}$ [78, 115]. La combinaison de la plupart des substituants les plus efficaces donna naissance au ligand pseudo-peptidique "CGP78850" [64], qui s'avéra 200 fois plus actif que Ac-pYIN-NH_2 [60]. Le squelette tri-peptidique lui-même fut, dans une démarche séparée, remplacé afin d'aboutir à un petit ligand phosphaté totalement non peptidique [116], mais il semble que des optimisations sur cette nouvelle base ne furent pas tentées.*

Extension du récepteur ciblé (INSERM/CNRS)

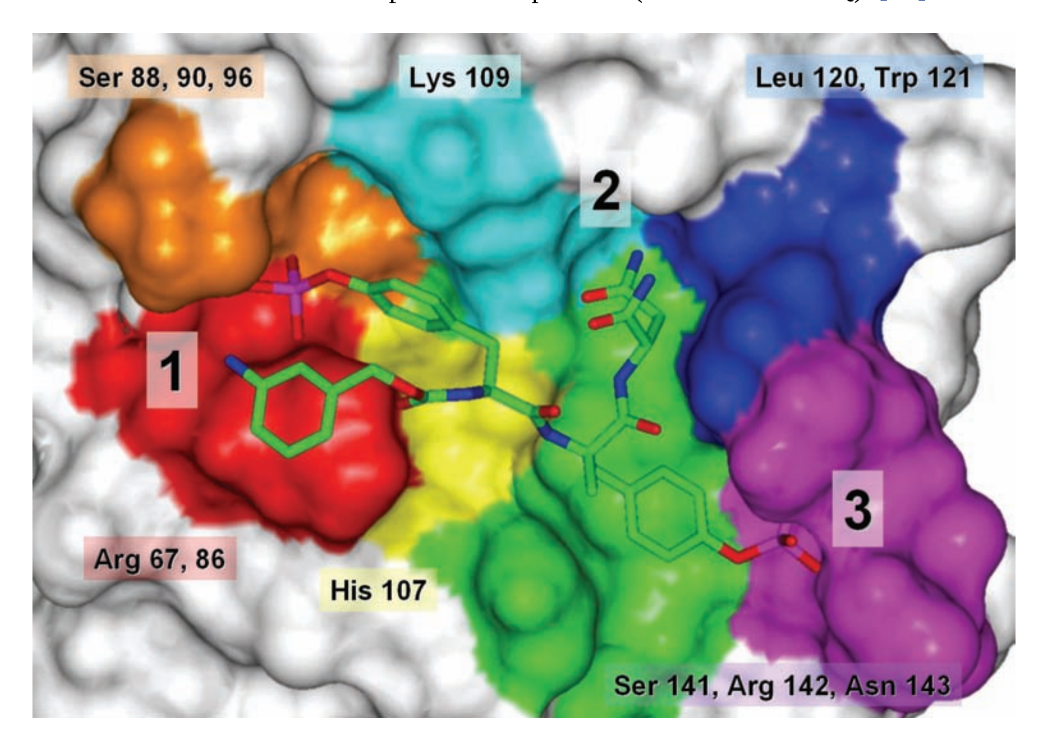
L'approche d'optimisation initiée par Novartis, consistant à substituer les différents groupes d'une séquence peptidique de référence (pYXN dans le cas de Grb2), est reprise par la plupart des équipes cherchant à mettre au point des inhibiteurs de domaines SH2. [52, 53, 83, 86, 109, 117-119] Une de ces équipes a permis, grâce à l'une de ces substitutions, de mettre en évidence qu'il était possible pour un ligand d'atteindre des zones du récepteur Grb2 SH2 non ciblées par les ligands connus jusqu'à présent, étendant ainsi la taille du récepteur.



Après avoir mis au point un ligand inhibant simultanément l'activité des deux domaines SH3 de Grb2 [120], l'équipe dirigée par Christiane Garbay s'intéressa à l'inhibition du domaine SH2 de Grb2, autre cible possible pour bloquer la voie de signalisation Ras-MAPK. Pour ce faire, ils prirent comme base un des pseudo-peptides développés par Novartis, mAz-pTyr-Ac_6c-Asn-NH_2 [88], dont les substituants aux positions -1 et +1 sont optimisés par rapport à la référence Ac-pYIN-NH_2, et tentèrent de déterminer un substituant encore plus efficace que Ac_6c_{+1} [79, 121]. D'autres travaux conduits à Novartis avaient indiqué qu'une séquence $pY_0pY_{+1}N_{+2}$ s'avérait 25 fois plus active sur le domaine SH2 de Src que la séquence pYYN obéissant au motif pYXN exigé par Grb2 SH2. [107] Le composé mAz-pTyr-pTyr-Asn-NH2 fut donc synthétisé, mais s'avéra deux fois moins actif que mAz-pTyr-Ac_6c-Asn-NH_2, indiquant que dans le cas de Grb2 le résidu $pTyr_{+1}$ n'était pas dans une conformation optimale. [79] Il fut donc α-méthylé dans le but de le stabiliser dans la conformation ß-turn spécifique aux ligands de Grb2 SH2. Le ligand résultant, mAz-pTyr-(α-Me)pTyr-Asn-NH_2 présenta cette fois-ci une activité dix fois plus importante (K_d = 30 ± 5 nM ; IC_{50} = 11 ± 1 nM) [79] ce qui en fait un des inhibiteurs les plus actifs pour Grb2 SH2 connus à ce jour.

* Ce ligand s'avéra en effet un peu moins actif que la référence Ac-pYIN-NH_2 (IC_{50} = 26 µM et 9 µM respectivement), ce qui n'en faisait pas à priori une base d'optimisation plus intéressante, d'autant plus que des modifications sur une telle base non peptidique sont notoirement plus difficiles à réaliser pour les expérimentateurs que des substitutions de groupes peptidiques.

De façon intéressante, des études par modélisation moléculaire de mAZ-pTyr-(α-Me)pTyr-Asn-NH_2 indiquèrent que si les résidus mAZ_{-1}, $pTyr_0$ et Asn_{+2} étaient liées dans les deux cavités connues du récepteur Grb2 SH2, (α-Me)$pTyr_{+1}$ interagissait dans une troisième cavité jusqu'alors non ciblée, composée des résidus S_{141} (Ser BG), R_{142} (Arg BG) et N_{143} (Asn BG). [79] Cette observation fut confirmée expérimentalement par la détermination de la structure RX du complexe correspondant (code PDB : 1JYQ). [68]



Structure expérimentale du ligand mAZ-pTyr-(α-Me)pTyr-Asn-NH_2 lié au récepteur de Grb2 SH2. Les principaux résidus actifs sont mentionnés ; on distingue clairement trois cavités, notées 1, 2 et 3.

Substituants à pTyr et cyclisation (Affymax, NCI)

Du moment qu'un résidu Asn_{+2} est présent dans la séquence, on peut supposer que la cyclisation du ligand n'est pas incompatible avec le maintien de la conformation *ß-turn* requise par Grb2 SH2. Une équipe du centre de recherche Affymax montra ainsi que la cyclisation d'une séquence peptidique C…pYXN…C par un pont disulfure, peut augmenter l'affinité pour Grb2 SH2, d'un ordre de grandeur dans certains cas. [96] Cela souligne le grand intérêt que constitue la stratégie de cyclisation afin de mettre au point des ligands plus efficaces de Grb2 SH2.

En se basant sur l'inhibiteur "CGP78850" de Novartis, Burke Jr. *et al.* (NCI), qui auparavant avaient déjà étudié les effets d'analogues de pTyr sur des inhibiteurs de SH2 [122, 123], orientèrent les optimisations ultérieures de l'inhibition de Grb2 SH2 dans deux directions distinctes : d'une part, l'emploi de divers substituants au groupe principal dérivé de pTyr [124-126], et d'autre part, la cyclisation du squelette peptidique afin de stabiliser la conformation *ß-turn* nécessaire à la reconnaissance du ligand [82, 127-131]. La combinaison de ces deux approches [132, 133] donna lieu à une classe de ligands cycliques non peptidiques efficaces [131, 134]. La structure RX de l'un d'entre eux [74] (PDB : 2AOA, 2AOB) indique une possibilité d'action simultanée sur deux domaines SH2*.

* Il peut s'agir d'artefacts de cristallisation, d'autant plus que pour les besoins de la RX le domaine SH2 a été isolé des deux domaines SH3 qui l'encadrent dans Grb2, et qui auraient pu constituer une gène stérique interdisant toute dimérisation SH2-ligand-SH2 de ce type. Cette dimérisation observée par RX souligne tout de même un caractère versatile du ligand, capable de se lier sous certaines conditions à Grb2 SH2 selon deux modes de liaison distincts.

Recherche de ligands actifs *in vivo* : contraintes

Une molécule active *in vitro* obéit à des contraintes de complémentarité géométrique et chimique par rapport à la cible biologique qu'elle vise. Une telle molécule ne sera pas nécessairement active *in vivo*, car pour cela elle doit obéir à trois contraintes supplémentaires qui sont en rapport avec le milieu biologique : (1) elle doit être *spécifique* à sa cible, (2) elle doit être *stable* chimiquement dans l'organisme, et (3) elle doit pouvoir se *diffuser* dans l'organisme jusqu'à la localisation du récepteur cible.

Contrainte de spécificité

Au récepteur cible peuvent correspondre dans l'organisme plusieurs homologues dont les fonctions biologiques sont différentes. Dans pareil cas l'inhibiteur se doit d'être spécifique afin d'être suffisamment actif sur la cible sans avoir à recourir à des dosages importants qui pourraient engendrer des effets secondaires indésirables.

Dans le cadre de la recherche de ligands actifs *in vivo*, la spécificité est ainsi une contrainte qui s'impose aussi bien à la cible qu'à l'inhibiteur. La spécificité d'un récepteur peut être estimée par des recherches bibliographiques et des techniques de base de la protéomique tels que les logiciels d'alignement de séquences. À ce niveau, les spécificités structurales du domaine SH2 de Grb2 contribuent grandement à son intérêt en tant que cible thérapeutique au sein de la voie de signalisation Ras-MAPK.

Estimer la spécificité d'un inhibiteur quelconque pour une cible donnée est, par contre une tâche bien plus délicate ; il s'agit d'une des raisons expliquant le taux d'échec élevé des protocoles de recherche de médicaments. Pour garantir qu'un ligand soit spécifique, il faudrait en effet le tester sur l'ensemble des biomolécules avec lesquelles il est susceptible d'interagir au sein d'un organisme, ce qui ne peut se faire qu'expérimentalement, en milieu clinique. Au niveau théorique, non seulement l'ensemble du protéome n'est pas connu, mais en plus l'espace connu est déjà trop vaste pour envisager son exploration systématique. On doit donc se limiter à l'emploi de méthodes statistiques sur des bases de molécules actives connues, ce qui est très limité, tant la portion de l'interactome couverte par ces molécules est restreinte.

Contrainte de stabilité

Dans l'organisme, les réactions de phosphorylation et de déphosphorylation sont courantes, ce qui pose problème au niveau des inhibiteurs de Grb2 SH2 disposant d'un résidu pTyr ou de tout autre résidu phosphaté : celui-ci est susceptible d'être déphosphorylé par les phosphatases ; le ligand perdrait alors une grande partie de son affinité pour Grb2 SH2.

De façon intéressante, l'étude par screening réalisée par Hart *et al.* [96] sur les séquences non phosphorylées de type $X_{-5}\ldots Y_0XN_{+2}\ldots X_{+10}$, si elle confirme que la déphosphorylation aboutit à une baisse importante de l'affinité du ligand (de un à trois ordres de grandeur), semble indiquer également que l'absence du groupe phosphate en position 0 a pour conséquence de déplacer la sélectivité sur les autres résidus du ligand, avec en particulier pour un Glu_{-2}, et de façon moins notable aux positions +4, +5 et +6. Cela pourrait traduire un mode de liaison légèrement différent des séquences peptidiques non phosphatées et souligne l'importance des groupes "distants" dans la mise au point de ligands pour Grb2 SH2 dans lesquels le groupe pTyr est substitué.

Contrainte d'accessibilité

Il existe des règles générales simples que doit respecter une molécule quelconque pour pouvoir être administrée à l'homme à la manière d'un médicament.[*] Il est évident que de telles règles doivent être considérées durant la recherche d'inhibiteurs pour une cible biologique donnée. L'identification de molécules non ”*drug like*”[†] efficaces *in vitro* peut toutefois s'avérer utile. En effet, il peut être possible, partant de ces molécules ”inutiles” comme référence, d'effectuer des modifications afin de leur conférer un caractère *drug-like* sans pour autant restreindre de façon significative leur efficacité.

Le respect de la contrainte d'accessibilité pour des inhibiteurs de Grb2 SH2 se heurte à une difficulté de taille : Grb2 est une protéine qui agit en milieu intra-cellulaire au sein de la voie de signalisation Ras-MAPK. Une molécule destinée à inhiber cette activité devra donc passer la barrière biologique que constitue la membrane cellulaire. Or tous les inhibiteurs connus de Grb2 SH2, y compris ceux pour lesquels le groupe pTyr a été substitué, sont hautement chargés négativement, ce qui leur interdit *a priori* le franchissement de la membrane. La mise au point d'inhibiteurs neutres ou faiblement chargés conservant une activité suffisante[‡] à partir des ligands connus de Grb2 SH2 constitue une tâche encore plus difficile que celle de la simple substitution de la phosphotyrosine.

Pourtant, il a été montré que le récepteur du domaine SH2 de SAP pouvait être lié à des inhibiteurs non seulement non phosphatés, mais aussi non chargés, inhibiteurs peptidiques qui plus est. En effet, SAP SH2 peut se lier avec une affinité (K_d = 0.6 μM) à SLAM autour du résidu Y_{281} lorsque celui-ci n'est pas phosphorylé.[§] [135, 136] L'analyse de la structure RX [137] ou RMN [138] du complexe formé indique que le mode de liaison diffère de celui des ligands typiques des domaines SH2 en général, car il implique la liaison dans trois cavités sur le récepteur et non plus deux. Des expériences de screening déterminèrent la séquence de référence comme étant (T/S)XX_0XX(V/I), ce qui indique que la sélectivité n'est pas déterminée par un résidu Y_0 ou pY_0 mais principalement aux positions -2 et +3. [138] Cela indique, au niveau de la séquence de référence TIY_0AQV sur SLAM, que la somme des interactions des résidus T_{-2} et V_{+3} de la séquence TIY_0AQV, est plus importante que celle du résidu Y_0.[**]

Ces travaux sur SAP SH2 n'ont pas (à notre connaissance) donné lieu à la mise au point d'un inhibiteur SH2 neutre et actif *in vivo*. Ils montrent toutefois que le ciblage de cavités ”lointaines” par des groupes neutres est susceptible, du moins pour certains domaines SH2, de donner lieu à des interactions qui pourraient, après optimisation, compenser celles du résidu pTyr de référence.

En ce qui concerne Grb2 SH2, deux cavités distinctes de celle du résidu pTyr et dans lesquelles des ligands peuvent se lier partiellement sont connues. [68] Cette caractéristique est encourageante dans l'optique de la mise au point de ligands moins chargés (et donc plus susceptibles de passer la barrière membranaire).

---

[*] Les règles de Lipinski, pour citer les plus connues, permettent d'exclure de façon relativement fiable des molécules qui ne pourraient pas être administrées oralement et ensuite se diffuser librement dans l'organisme.

[†] Anglicisme pratique désignant le caractère ”médicamenteux” d'une molécule sur le plan structural, ce qui a trait entre autres au respect ou non des règles de Lipinski.

[‡] C'est-à-dire, un million de fois plus efficaces que la séquence peptidique de référence pYVNV, chargée doublement négativement et qui contient le résidu pTyr.

[§] L'affinité augmente après phosphorylation, mais seulement d'un facteur 3, à comparer aux quatres ordres de grandeur observés en ce qui concerne la plupart des domaines SH2.

[**] Lorsque le résidu Y_0 est phosphorylé, on peut supposer que son interaction avec SAP SH2 devient alors comparable à celle des résidus voisins. Cela contraste avec le cas de la plupart des ligands de domaines SH2 ciblant les deux cavités du récepteur, dans lesquels le résidu pY_0 semble irremplaçable.

Recherche de ligands actifs *in vivo* : stratégies pour Grb2 SH2

Le développement de ligands actifs *in vivo* pour Grb2 SH2 peut se faire soit à partir de ligands actifs *in vitro*, soit en tentant de mettre au point des bases structurales nouvelles en intégrant dès le début les contraintes spécifiques de stabilité et d'accessibilité.

Modification de ligands existants

Une modification intéressante de la phosphotyrosine a été proposée par Gay *et al.* afin de permettre la pénétration cellulaire d'un ligand (”CGP78850”, dont nous avons parlé précédemment) comprenant déjà un analogue de phosphotyrosine plus résistant aux phosphatases. Afin de permettre la pénétration de la membrane cellulaire, ce groupe est estérifié, ce qui donne le composé ”CGP85793”. Celui-ci, après pénétration cellulaire, sera hydrolysé par les estérases intra-cellulaires [139], relâchant ”CGP78850” qui peut alors inhiber Grb2 SH2. [60, 64]

D'une façon générale, une démarche couramment entreprise afin d'augmenter l'accessibilité *in vivo* des ligands à base de phosphotyrosine consiste à employer un substituant au groupe pTyr, afin de contourner les problèmes d'accessibilité et de stabilité qui lui sont liés. [77, 124, 125, 140] (*voir tableau page suivante*) Celui-ci peut être un autre groupe chargé à base de phosphore (en particulier Pmp et ses dérivés), un groupe acide, souvent à base carboxyl [126, 141], ou bien un groupe neutre électriquement, typiquement Tyr. À noter que sur aucun ligand connu de Grb2 SH2, tout comme pour Src SH2 [77], la substitution du groupe $pTyr_0$ n'a donné lieu à un gain d'affinité significatif.[*] [89, 119, 126] La limite de cette approche réside dans le fait que l'impact sur l'affinité d'un substituant donné (en particulier sur $pTyr_0$) dépend grandement de la base structurale sur laquelle la substitution est effectuée. Par exemple, Pmp peut s'avérer, suivant la plateforme sur laquelle il se substitue à $pTyr_0$, soit équivalent soit largement défavorable. [19]

Nouvelles bases structurales

Une autre approche pour la détermination de ligands non phosphorylés pour Grb2 SH2 consiste à reprendre les méthodes qui ont permis par optimisations successives de déterminer des ligands phosphorylés efficaces : sachant que les séquences peptidiques $pY_0X_{+1}N_{+2}$ sont préférées, il peut s'avérer intéressant d'effectuer des screenings de séquences $Y_0X_{+1}N_{+2}$. Sur de telles séquences, on remarque alors que l'absence du groupe phosphate en position 0 entraîne une forte sélection pour un résidu Glu_{-2}, qui compenserait ainsi l'absence de charge négative sur la position 0. [96]

[*] On observe dans le meilleur des cas (pour des ligands peu ou moyennement actifs, et avec des substituants doublement chargés tels que Pmp, pmF et FOMT) une activité légèrement meilleure, sans que ce gain soit significatif au regard de la précision intrinsèque des protocoles expérimentaux de mesure d'affinité. La plupart du temps, on observe une perte d'affinité d'un ou plusieurs ordres de grandeur, d'autant plus importante que la charge négative portée par le groupe phosphate de pTyr est diminuée.

Roller *et al.* (NCI) réussirent ainsi à déterminer la séquence peptidique cyclique (nommée G1TE) $CELY_0E_{+1}N_{+2}VGMYC$, qui présente une affinité intéressante et pour laquelle, de façon remarquable, les deux résidus acides Glu_{-2} et Glu_{+1} semblent compenser l'absence de groupe phosphate en position 0. [95] Des expériences de mutagenèse sur cette base montrèrent que chaque résidu de la séquence de G1TE, excepté Gly_{+4}, a un rôle dans l'affinité pour Grb2 SH2, et confirmèrent l'importance particulière des résidus Glu_{-2} et Asn_{+2}. [95, 142, 143] De façon intéressante, la structure RMN du complexe de G1TE avec Grb2 SH2 indiqua que G1TE ne forme aucune liaison hydrogène avec Grb2 SH2, les chaînes latérales des peptides étant toutes tournées à l'extérieur du récepteur de Grb2 SH2. Ainsi, on peut supposer d'une part que l'affinité de G1TE pour Grb2 SH2 provient de la conformation adoptée par le cycle, qui serait particulièrement favorable. [144] D'autre part, G1TE constitue une base intéressante pour la mise au point d'inhibiteurs non phosphorylés de Grb2 SH2, au même titre que la séquence Ac-pYIN-NH_2 utilisée initialement par Novartis pour la mise au point d'inhibiteurs phosphorylés. Différentes substitutions donnèrent en effet lieu à des composés certes actifs [145-151], mais dont la charge totale reste de -2.

pTyr	phosphotyrosine	R-CH_2-Ø-O-PO_3^{2-}	
Tyr	tyrosine	R-CH_2-Ø-OH	
-	phosphonotyrosine	R-CH_2-Ø-PO_3^{2-}	
-	phosphinate isosteres	R-CH_2-Ø-CH_2-PO_2^--X	[111, 112]
Pmp	phosphonomethyl-phenylalanine	R-CH_2-Ø-CH_2-PO_3^{2-}	[64, 123, 147, 152-155]
F_2Pmp	phosphonodifluoromethyl-phenylalanine	R-CH_2-Ø-CF_2-PO_3^{2-}	[123, 153, 156, 157]
cmF	carboxymethyl-phenylalanine	R-CH_2-Ø-CH_2-COO^-	[117, 141]
F_2cmF	carboxydifluoromethyl-phenylalanine	R-CH_2-Ø-CF_2-COO^-	[141, 158]
pmF	P-malonyl-phenylalanine	R-CH_2-Ø-CH-$(COO^-)_2$	[132, 159]
Gla	acide γ-carboxyglutamique	R-CH_2-CH-$(COO^-)_2$	[143, 146]
OMT	O-malonyl-tyrosine	R-CH_2-Ø-O-CH-$(COO^-)_2$	[160, 161]
FOMT	fluoro-O-malonyl-tyrosine	R-CH_2-Ø-O-CF-$(COO^-)_2$	[162]
cmT	carboxymethyl-tyrosine	R-CH_2-Ø-O-CH_2-COO^-	[163]

Substituants à pTyr employés dans les inhibiteurs de Grb2 SH2
De gauche à droite : abréviation usuelle, nom chimique, formule, références

Page suivante :

Sélection d'inhibiteurs de Grb2 SH2 – Références :

1 [78, 113], 2 [64], 3 [64], 4 [71], 5 [68, 79], 6 [82, 134], 7 [74], 8 [95, 144], 9 [146]

Abréviations :

Ac_6c *= acide 1-aminocyclohexyl-carboxylique ;* ***hNp*** *= 3-(2-hydroxynaphthalen-1-yl)-propyle*
mAZ *= m-aminobenzylo-carbonyle ;* ***(S,R)Achec*** *= acide cis-2-amino-cyclohex-3-yl-carboxylique*
naph *= D-2-naphthyl-alanine ;* ***meF*** *= methyl-phenylalanine*
tBuAla *= tertiobutyl-alanine ;* ***AmF*** *= amino-phenylalanine*

1 (référence)

pY_0, N_{+2}, I_{+1}

2 Novartis CGP78850

Pmp_0, N_{+2}, hNp_{+3}, Ac_6c_{+1}

3 Novartis CGP85793

(protected) Pmp_0, N_{+2}, hNp_{+3}, Ac_6c_{+1}

4 Novartis PDB : 1CJ1

pY_0, (S,R)$Achec_{+2}$, Ac_6c_{+1}, mAz_{-1}

6 Burke *et al.* (NCI) PDB : 1X0N

$naph_{-1}$, N_{+2}, Ac_6c_{+1}, Pmp_0

8 Roller *et al.* (NCI) G1TE

N_{+2}, V_{+3}, G_{+4}, M_{+5}, Y_{+6}, E_{+1}, C, Y_0, L_{-1}, E_{-2}

5 Garbay *et al.* (INSERM/CNRS) PDB : 1JYQ

pY_0, N_{+2}, Me_α, pY_{+1}, mAz_{-1}

7 Burke *et al.* (NCI) PDB: 2AOA, 2AOB

$naph_{-1}$, N_{+2}, Ac_6c_{+1}, Pmf_0, D_α

9 Roller *et al.* (NCI)

N_{+2}, V_{+3}, G_{+4}, $tBuAla_{+5}$, AmF_{+6}, Ac_6c_{+1}, C, Y_0, MeF_{-1}, Gla_{-2}

Références bibliographiques

1. Vidal M., Gigoux V. and Garbay C. SH2 and SH3 domains as targets for anti-proliferative agents. *Critical Reviews in Oncology / Hematology* **40** (2001) 175-186.

2. Robbins D.J., Cheng M., Zhen E., Vanderbilt C.A., Feig L.A. and Cobb M.H. Evidence for a Ras-dependent extracellular signal-regulated protein kinase (ERK) cascade. *Proceedings of the National Academy of Sciences USA* **89** (1992) 6924-6928.

3. Schlessinger J. How receptor tyrosine kinases activate Ras. *Trends in Biochemical Sciences* **18** (1993) 273-275.

4. Khosravi-Far R. and Der C.J. The Ras signal transduction pathway. *Cancer and Metastasis Reviews* **13**, issue 1 (1994) 67-89.

5. Vojtek A.B. and Der C.J. Increasing complexity of the Ras signaling pathway. *Journal of Biological Chemistry* **273**, issue 32 (1998) 19925-19928.

6. Reuter C.W.M., Morgan M.A. and Bergmann L. Targeting the Ras signaling pathway: a rational, mechanism-based treatment for hematologic malignancies? *Blood* **96**, issue 5 (2000) 1655-1669.

7. Olsson A.-K. Ras-MAPK signaling in differentiating SH-SY5Y human neuroblastoma cells, in *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* (2000) Academic press / Acta Universitatis Upsaliensis, Uppsala, p. 66.

8. Alberola-Ila J. and Hernández-Hoyos G. The Ras/MAPK cascade and the control of positive selection. *Immunological Reviews* **191** (2003) 79-96.

9. Brambilla R., Gnesutta N., Minichiello L., White G., Roylance A.J., Herron C.E., Ramsey M., Wolfer D.P., Cestari V., Rossi-Arnaud C., Grant S.G., Chapman P.F., Lipp H.P., Sturani E. and Klein R. A role for the Ras signalling pathway in synaptic transmission and long-term memory. *Nature* **390** (1997) 281-286.

10. Bos J.L. Ras oncogenes in human cancer: a review. *Cancer Research* **49** (1989) 4682-4689.

11. Davies H., Bignell G.R., Cox C., Stephens P., Edkins S., Clegg S., Teague J., Woffendin H., Garnett M.J., Bottomley W., Davis N., Dicks E., Ewing R., Floyd Y., Gray K., Hall S., Hawes R., Hughes J., Kosmidou V., Menzies A., Mould C., Parker A., Stevens C., Watt S., Hooper S., Wilson R., Jayatilake H., Gusterson B.A., Cooper C., Shipley J., Hargrave D., Pritchard-Jones K., Maitland N., Chenevix-Trench G., Riggins G.J., Bigner D.D., Palmieri G., Cossu A., Flanagan A., Nicholson A., Ho J.W.C., Leung S.Y., Yuen S.T., Weber B.L., Seigler H.F., Darrow T.L., Paterson H., Marais R., Marshall C.J., Wooster R., Stratton M.R. and Futreal P.A. Mutations of the BRAF gene in human cancer. *Nature* **417** (2002) 949-954.

12. Garbay C., Liu W.-Q., Vidal M. and Roques B.P. Inhibitors of Ras signal transduction as antitumor agents. *Biochemical Pharmacology* **60** (2000) 1165-1169.

13. Liu W.-Q., Vidal M., Mathé C., Périgaud C. and Garbay C. Inhibition of the Ras-dependent mitogenic pathway by phosphopeptide prodrugs with antiproliferative properties. *Bioorganic & Medicinal Chemistry Letters* **10** (2000) 669-672.

14. Barbacid M. Ras genes. *Annual Review of Biochemistry* **56** (1987) 779-827.

15. Warne P.H., Viciana P.R. and Downward J. Direct interaction of Ras and the amino-terminal region of Raf-1 in vitro. *Nature* **364** (1993) 352-355.

16. Zhang X.F., Settleman J., Kyriakis J.M., Takeuchi-Suzuki E., Elledge S.J., Marshall M.S., Bruder J.T., Rapp U.R. and Avruch J. Normal and oncogenic p21ras proteins bind to the amino-terminal regulatory domain of c-Raf-1. *Nature* **364** (1993) 308-313.

17. Basu T., Warne P.H. and Downward J. Role of Shc in the activation of Ras in response to epidermal growth factor and nerve growth factor. *Oncogene* **9** (1994) 3483-3491.

18. Skolnik E.Y., Batzer A., Li N., Lee C.-H., Lowenstein E., Maohammadi M., Margolis B. and Schlessinger J. The function of GRB2 in linking the insulin receptor to Ras signaling pathways. *Science* **260**, issue 5116 (1993) 1953-1955.

19. Machida K. and Mayer B.J. The SH2 domain: versating signaling module and pharmaceutical target. *Biochimica et Biophysica Acta* **1747** (2005) 1-25.

20. Sadowski I., Stone J.C. and Pawson T. A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130(gag-fps). *Molecular and Cellular Biology* **6**, issue 12 (1986) 4396-4408.

21. Russel R.B., Breed J. and Barton G.J. Conservation analysis and structure prediction of the SH2 family of phosphotyrosine binding domains. *FEBS Letters* **304** (1992).

22. Pawson T. and Nash P. Assemby of cell regulatory systems through protein interaction domains. *Science* **300** (2003) 445-452.

23. Eck M.J., Shoelson S.E. and Harrison S.C. Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. *Nature* **362** (1993) 87-91.

24. Songyang Z., Shoelson S.E., Chaudhuri M., Gish G., Pawson T., Haser W.G., King F., Roberts T., Ratnofsky S., Lechleider R.J., Neel B.G., Birge R.B., Fajardo J.E., Chou M.M., Hanafusa H., Schaffhausen B. and Cantley L.C. SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**, issue 5 (1993) 767-778.

25. Marangere L.E.M. and Pawson T. Structure and function of SH2 domains. *Journal of Cell Science* **Suppl. 18** (1994).

26. Pawson T. and Gish G.D. SH2 and SH3 domains: from structure to function. *Cell* **71** (1992) 359-362.

27. Pawson T. Protein modules and signalling networks. *Nature* **373** (1995) 573-580.

28. Moran M.F., Koch C.A., Anderson D., Ellis C., England L., Martin G.S. and Pawson T. Src homology region 2 domains direct protein-protein interactions in signal transduction. *Proceedings of the National Academy of Sciences USA* **87** (1990) 8622-8626.

29. Pawson T., Olivier P., Rozakis-Adcock M., McGlade J. and Henkemeyer M. Proteins with SH2 and SH3 domains couple receptor tyrosine kinases to intracellular signalling pathways. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences* **340**, issue 1293 (1993) 279-285.

30. Smithgall T.E. SH2 and SH3 domains: Potential targets for anti-cancer drug design. *Journal of Pharmacological and Toxicological Methods* **34** (1995) 125-132.

31. Waksman G., Kumaran S. and Lubman O.Y. SH2 domains: role, structure and implications for molecular medicine. *Expert Reviews in Molecular Medicine* **6**, issue 3 (2004).

32. Matuoka K., Shibata M., Yamakawa A. and Takenawa T. Cloning of ASH, a ubiquitous protein composed of one Src homology region (SH) 2 and two SH3 domains, from human and rat cDNA libraries. *Proceedings of the National Academy of Sciences USA* **89** (1992) 9015-9019.

33. Lowenstein E.J., Daly R.J., Batzer A.G., Li W., Margolis B., Lammers R., Ullrich A., Skolnik E.Y., Bar-Sagi D. and Schlessinger J. The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. *Cell* **70**, issue 3 (1992) 431-442.

34. Chardin P., Cussac D., Maignan S. and Ducruix A. The Grb2 adaptor. *FEBS Letters* **369** (1995) 47-51.

35. Maignan S., Guilloteau J.P., Fromage N., Arnoux B., Becquart J. and Ducruix A. Crystal structure of the mammalian Grb2 adaptor. *Science* **268**, issue 5208 (1995) 291-293.

36. Salcini A.E., McGlade J., Pelicci G., Nicoletti I., Pawson T. and Pelicci P.G. Formation of Shc-Grb2 complexes is necessary to induce neoplastic transformation by overexpression of Shc proteins. *Oncogene* **9**, issue 10 (1994) 2827-2836.

37. Xie Y., Li K. and Hung M.C. Tyrosine phosphorylation of Shc proteins and formation of Shc/Grb2 complex correlate to the transformation of NIH3T3 cells mediated by the point-mutation activated neu. *Oncogene* **10**, issue 12 (1995) 2409-2413.

38. Buday L. Membrane-targeting of signalling molecules by SH2/SH3 domain-containing adaptor proteins. *Biochimica et Biophysica Acta* **1422** (1999) 187-204.

39. Tari A.M. and Lopez Berestein G. Grb2: a pivotal protein in signal transduction. *Seminars in Oncology* **28**, issue 5 suppl 16 (2001) 142-147.

40. Li N., Batzer A., Daly R., Yajnik V., Skolnik E., Chardin P., Bar-Sagi D., Margolis B. and Schlessinger J. Guanine-nucleotide-releasing factor hSos1 binds to Grb2 and links receptor tyrosine kinases to Ras signalling. *Nature* **363** (1993) 85-88.

41. Rozakis-Adcock M., Fernley R., Wade J., Pawson T. and Bowtell D. The SH2 and SH3 domains of mammalian Grb2 couple the EGF receptor to the Ras activator mSos1. *Nature* **363** (1993) 83-85.

42. Rozakis-Adcock M., McGlade J., Mbamalu G., Pelicci G., Daly R., Li W., Batzer A., Thomas S., Brugge J., Pelicci P.G., Schlessinger J. and Pawson T. Association of the Shc and Grb2/Sem5 SH2-containing proteins is implicated in activation of the Ras pathway by tyrosine kinases. *Nature* **360** (1992) 689-692.

43. Skolnik E.Y., Lee C.-H., Batzer A.G., Vicentini L.M., Zhou M., Daly R.J., Myers Jr. M.J., Backer J.M., Ullrich A., White M.F. and Schlessinger J. The SH2/SH3 domain-containing protein Grb2 interacts with tyrosine-phosphorylated IRS1 and Shc: implications for insulin control of Ras signaling. *EMBO Journal* **12** (1993) 1929-1936.

44. Schlaepfer D.D., Hanks S.K., Hunter T. and van der Geer P. Integrin-mediated signal transduction linked to Ras pathway by Grb2 binding to focal adhesion kinase. *Nature* **372** (1994) 786-791.

45. Atabey N., Gao Y., Yao Z.-J., Breckenridge D., Soon L., Soriano J.V., Burke Jr. T.R. and Bottaro D.P. Potent blockade of hepatocyte growth factor-stimulated cell motility, matrix invasion and branching morphogenesis by antagonists of Grb2 Src homology 2 domain interactions. *Journal of Biological Chemistry* **276**, issue 17 (2001) 14308-14314.

46. Neel B.G. and Tonks N.K. Protein tyrosine phosphatases in signal transduction. *Current Opinion in Cellular Biology* **9**, issue 2 (1997) 193-204.

47. Lawrence D.S. and Niu J. Protein kinase inhibitors: the tyrosine-specific protein kinases. *Pharmacology & Therapeutics* **77**, issue 2 (1998) 81-114.

48. Levitzki A. Protein tyrosine kinase inhibitors as novel therapeutic agents. *Pharmacology & Therapeutics* **82**, issue 2-3 (1999) 231-239.

49. Baselga J. Targeting tyrosine kinases in cancer: the second wave. *Science* **312**, issue 5777 (2006) 1175-1178.

50. Force T., Kuida K., Namchuk M., Parang K. and Kyriakis J.M. Inhibitors of protein kinase signaling pathways: Emerging therapies for cardiovascular disease. *Circulation* **109** (2004) 1196-1205.

51. Botfield M.C. and Green J. SH2 and SH3 domains: choreographers of multiple signaling pathways. *Annual Reports in Medicinal Chemistry* **30** (1995) 227-237.

52. Sawyer T.K. Src homology-2 domains: Structure, mechanisms, and drug discovery. *Biopolymers (Peptide Science)* **47** (1998) 243-261.

53. Cody W.L., Lin Z., Panek R.L., Rose D.W. and Rubin J.R. Progress in the development of inhibitors of SH2 domains. *Current Pharmaceutical Design* **6** (2000) 59-98.

54. Madema R.H. and Bos J.L. The role of p21-ras in receptor tyrosine kinase signaling. *Critical Reviews in Oncology / Hematology* **4** (1993) 615-661.

55. Tari A.M., Arlinghaus R. and Lopez Berestein G. Inhibition of Grb2 and Crkl proteins results in growth inhibition of Philadelphia chromosome positive leukemic cells. *Biochemical and Biophysical Research Communications* **235** (1997) 383-388.

56. Daly R.J., Binder M.D. and Sutherland R.L. Overexpression of the Grb2 gene in human breast cancer cell lines. *Oncogene* **9**, issue 9 (1994) 2723-2727.

57. Sastry L., Cao T. and King C.R. Multiple Grb2-protein complexes in human cancer cells. *International Journal of Cancer* **70** (1997) 208-213.

58. Kairouz R. and Daly R.J. Modulation of tyrosine kinase signalling in human breast cancer through altered expression of signalling intermediates. *Breast Cancer Research and Treatment* **2** (2000) 197-202.

59. Pendergast A.M., Quilliam L.A., Cripe L.D., Bassing C.H., Dai Z., Li N., Batzer A., Rabun K.M., Der C.J., Schlessinger J. and others BCR-ABL-induced oncogenesis is mediated by direct interaction with the SH2 domain of the GRB-2 adaptor protein. *Cell* **75**, issue 1 (1993) 175-185.

60. Gay B., Suarez S., Weber C., Rahuel J., Fabbro D., Furet P., Caravatti G. and Schoepfer J. Effect of potent and selective inhibitors of the Grb2 SH2 domain on cell motility. *Journal of Biological Chemistry* **274**, issue 33 (1999) 23311-23315.

61. Soriano J.V., Liu N., Gao Y., Yao Z.-J., Ishibashi T., Underhill C., Burke Jr. T.R. and Bottaro D.P. Inhibition of angiogenesis by growth factor receptor bound protein 2-Src homology 2 domain binding antagonists. *Molecular Cancer Therapeutics* **3**, issue 10 (2004) 1289-1299.

62. Brugge J.S. New intracellular targets for therapeutic drug design. *Science* **260** (1993) 918-919.

63. Boutin J.A. Tyrosine proteine kinase inhibition and cancer. *International Journal of Biochemistry* **26**, issue 10-11 (1994) 1203-1226.

64. Gay B., Suarez S., Caravatti G., Furet P., Meyer T. and Schoepfer J. Selective Grb2 SH2 inhibitors as anti-Ras therapy. *International Journal of Cancer* **83** (1999) 235-241.

65. Guilloteau J.P., Fromage N., Ries-Kautt M., Reboul S., Bocquet D., Dubois H., Faucher D., Colonna C., Ducruix A. and Becquart J. Purification, stabilization, and crystallization of a modular protein: Grb2. *Proteins: Structure, Function, and Genetics* **25**, issue 1 (1996) 112-119.

66. Thornton K.H., Mueller W.T., McConnell P., Zhu G., Saltiel A.R. and Thanabal V. Nuclear magnetic resonance solution structure of the growth factor receptor-bound protein 2 Src homology 2 domain. *Biochemistry* **35** (1996) 11852-11864.

67. Senior M.M., Frederick A.F., Black S., Murgolo N.J., Perkins L.M., Wilson O., Snow M.E. and Wang Y.-S. The three-dimensional solution structure of the Src homology domain-2 of the growth factor receptor-bound protein-2. *Journal of Biomolecular NMR* **11** (1998) 153-164.

68. Nioche P., Liu W.-Q., Broutin I., Charbonnier F., Latreille M.-T., Vidal M., Roques B., Garbay C. and Ducruix A. Crystal structures of the SH2 domain of Grb2: Highlight on the binding of a new high-affinity inhibitor. *Journal of Molecular Biology* **315** (2002) 1167-1177.

69. Rahuel J., García-Echeverría C., Furet P., Strauss A., Caravatti G., Fretz H., Schoepfer J. and Gay B. Structural basis for the high affinity of amino-aromatic SH2 phosphopeptide ligands. *Journal of Molecular Biology* **279** (1998) 1013-1022.

70. Ettmayer P., France D., Gounarides G., Jarosinski M., Martin M.-S., Rondeau J.-M., Sabio M., Topiol S., Weidmann B., Zurini M. and Bair K.W. Structural and conformational requirements for high-affinity binding to the SH2 domain of Grb2. *Journal of Medicinal Chemistry* **42** (1999) 971-980.

71. Furet P., García-Echeverría C., Gay B., Schoepfer J., Zeller M. and Rahuel J. Structure-based design, synthesis, and X-ray crystallography of a high-affinity antagonist of the Grb2-SH2 domain containing an asparagine mimetic. *Journal of Medicinal Chemistry* **42** (1999) 2358-2363.

72. Ogura K., Tsuchiya S., Terasawa H., Yuzawa S., Hatanaka H., Mandiyan V., Schlessinger J. and Inagaki F. Solution structure of the SH2 domain of Grb2 complexed with the Shc-derived phosphotyrosine-containing peptide. *Journal of Molecular Biology* **289** (1999) 439-445.

73. Schiering N., Casale E., Caccia P., Giordano P. and Battistini C. Dimer formation through domain swapping in the crystal structure of the Grb2-SH2-Ac-pYVNV complex. *Biochemistry* **39** (2000) 13376-13382.

74. Phan J., Shi Z.-D., Burke Jr. T.R. and Waugh D.S. Crystal structures of a high-affinity macrocyclic peptide mimetic in complex with the Grb2 SH2 domain. *Journal of Molecular Biology* **353** (2005) 104-115.

75. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. and Bourne P.E. The Protein Data Bank. *Nucleic Acids Research* **28**, issue 1 (2000) 235-242.

76. Smith Schmidt T. Banking on structures. *BioIT World* **1**, issue 8 (2002).

77. Gilmer T., Rodriguez M., Jordan S., Crosby R., Alligood K., Green M., Kimery M., Wagner C., Kinder D., Charifson P., Hassell A.M., Willard D., Luther M., Runsak D., Sternbach D.D., Mehrotra M., Peel M., Shampine L., Davis R., Robbins J., Patel I.R., Kassel D., Buckhart W., Moyer M., Bradshaw T. and Berman J. Peptide inhibitors of src SH3-SH2-phosphoprotein interactions. *Journal of Biological Chemistry* **269**, issue 50 (1994) 31711-31719.

78. Furet P., Gay B., Caravatti G., García-Echeverría C., Rahuel J., Schoepfer J. and Fretz H. Structure-based design and synthesis of high affinity tripeptide ligands of the Grb2-SH2 domain. *Journal of Medicinal Chemistry* **41** (1998) 3442-3449.

79. Liu W.-Q., Vidal M., Gresh N., Roques B.P. and Garbay C. Small peptides containing phosphotyrosine and adjacent αMe-phosphotyrosine or its mimetics as highly potent inhibitors of Grb2 SH2 domain. *Journal of Medicinal Chemistry* **42** (1999) 3737-3741.

80. Burke Jr. T.R., Yao Z.-J., Gao Y., Wu J.X., Zhu X., Luo J.H., Guo R. and Yang D. N-Terminal carboxyl and tetrazole-containing amides as adjuvants to Grb2 SH2 domain ligand binding. *Bioorganic & Medicinal Chemistry Letters* **9** (2001) 1439-1445.

81. Wei C.-Q., Li B., Guo R., Yang D. and Burke Jr. T.R. Development of a phosphatase-stable phosphotyrosyl mimetic suitably protected for the synthesis of high-affinity Grb2 SH2 domain-binding ligands. *Bioorganic & Medicinal Chemistry Letters* **12** (2002) 2781-2784.

82. Wei C.-Q., Gao Y., Lee K., Guo R., Li B., Zhang M., Yang D. and Burke Jr. T.R. Macrocyclization in the design of Grb2 SH2 domain-binding ligands exhibiting high potency in whole-cell systems. *Journal of Medicinal Chemistry* **46** (2003) 244-254.

83. Lung F.-D.T. and Tsai J.-Y. Grb2 SH2 domain-binding peptide analogs as potential anticancer agents. *Peptide Science* **71** (2003) 132-140.

84. Gram H., Schmitz R., Zuber J.-F. and Baumann G. Identification of phosphopeptide ligands for the Src-homology 2 (SH2) domain of Grb2 by phage display. *European Journal of Biochemistry* **246** (1997) 633-637.

85. Lung F.-D.T., Tsai J.-Y., Wei S.-Y., Cheng J.-W., Chen C., Li P. and Roller P.P. Novel peptide inhibitors for Grb2 SH2 domain and their detection by surface plasmon resonance. *Journal of Peptide Research* **60** (2002) 143-149.

86. Shi Z.-D., Karki R.G., Oishi S., Worthy K.M., Bindu L.K., Dharmawardana P.G., Nicklaus M.C., Bottaro D.P., Fisher R.J. and Burke Jr. T.R. Utilization of a nitrobenzoxadianole (NBD) fluorophore in the design of a Grb2 SH2 domain-binding peptide mimetic. *Bioorganic & Medicinal Chemistry Letters* **15** (2005) 1385-1388.

87. Songyang Z., Shoelson S.E., McGlade J., Olivier P., Pawson T., Bustelo X.R., Barbacid M., Sabe H., Hanafusa H., Yi T., Ren R., Baltimore D., Ratnofsky S., Feldman R.A. and Cantley L.C. Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Molecular and Cellular Biology* **14**, issue 4 (1994) 2777-2785.

88. García-Echeverría C., Furet P., Gay B., Fretz H., Rahuel J., Schoepfer J. and Caravatti G. Potent antagonists of the SH2 domain of Grb2: Optimization of the X+1 position of 3-amino-Z-Tyr(PO_3H_2)-X^{+1}-Asn-NH_2. *Journal of Medicinal Chemistry* **41**, issue 11 (1998) 1741-1744.

89. Fretz H., Furet P., García-Echeverría C., Rahuel J. and Schoepfer J. Structure-based design of compounds inhibiting Grb2-SH2 mediated protein-protein interactions in signal transduction pathways. *Current Pharmaceutical Design* **6** (2000) 1777-1796.

90. Shakespeare W.C. SH2 domain inhibition: a problem solved? *Current Opinion in Chemical Biology* **5** (2001) 409-415.

91. Lemmon M.A., Ladbury J.E., Mandiyan V., Zhou M. and Schlessinger J. Independent binding of peptide ligands to the SH2 and SH3 domains of Grb2. *Journal of Biological Chemistry* **269**, issue 50 (1994) 31653-31658.

92. Cantley L.C., Auger K.R., Carpenter C., Duckworth B., Graziani A., Kapeller R. and Soltoff S. Oncogenes and signal transduction. *Cell* **65**, issue 5 (1991) 914.

93. Ward C.W., Gough K.H., Rashke M., Wan S.S., Tribbick G. and Wang J.-X. Systematic mapping of potential binding sites for Shc and Grb2 SH2 domains on Insuline Receptor Substrate-1 and the receptors for Insulin, Epidermal Growth Factor, Platelet-derived Growth Factor, and Fibroblast Growth Factor. *Journal of Biological Chemistry* **271**, issue 10 (1996) 5603-5609.

94. Müller K., Gombert F.O., Manning U., Grossmüller F., Graff P., Zaegel H., Zuber J.-F., Freuler F., Tschopp C. and Baumann G. Rapid identification of phosphopeptide ligands for SH2 domains. *Journal of Biological Chemistry* **271**, issue 28 (1996) 16500-16505.

95. Oligino L., Lung F.-D.T., Sastry L., Bigelow J., Cao T., Curran M., Burke Jr. T.R., Wang S., Krag D., Roller P.P. and King C.R. Nonphosphorylated peptide ligands for the Grb2 Src homology 2 domain. *Journal of Biological Chemistry* **272**, issue 46 (1997) 29046-29052.

96. Hart C.P., Martin J.E., Reed M.A., Keval A.A., Pustelnik M.J., Northrop J.P., Patel D.V. and Grove J.R. Potent inhibitory ligands of the Grb2 SH2 domain from recombinant peptide libraries. *Cellular Signalling* **11**, issue 6 (1999) 453-464.

97. McNemar C., Snow M.E., Windsor W.T., Prongay A., Mui P., Zhang R., Durkin C., Le H.V. and Weber P.C. Thermodynamic and structural analysis of phosphotyrosine polypeptide binding to Grb2-SH2. *Biochemistry* **36** (1997) 10006-10014.

98. Suenaga A., Hatakeyama M., Ichikawa M., Yu X., Futatsugi N., Narumi T., Fukui K., Terada T., Taiji M., Shirouzu M., Yokoyama S. and Konagaya A. Molecular dynamics, free energy, and SPR analyses of the interactions between the SH2 domain of Grb2 and ErbB phosphotyrosyl peptides. *Biochemistry* **42** (2003) 5195-5200.

99. Kessels H.W.H.G., Ward A.C. and Schumacher T.N.M. Specificity and affinity motifs for Grb2 SH2-ligand interactions. *Proceedings of the National Academy of Sciences USA* **99**, issue 13 (2002) 8524-8529.

100. Broadshaw J.M., Grucza R.A., Ladbury J.E. and Waksman G. Probing the “two-pronged plug two-holed socket” model for the mechanism of binding of the Src SH2 domain to phosphotyrosyl peptides: A thermodynamic study. *Biochemistry* **37** (1998) 9083-9090.

101. Booker G.W., Breeze A.L., Downing A.K., Panayotou G., Gout I., Waterfield M.D. and Campbell I.D. Structure of an SH2 domain of the p85α subunit of phosphatidylinositol-3-OH kinase. *Nature* **358** (1992) 684-687.

102. Overduin M., Mayer B., Rios C.B., Baltimore D. and Cowburn D. Secondary structure of Src homology 2 domain of c-Abl by heteronuclear NMR spectroscopy in solution. *Proceedings of the National Academy of Sciences USA* **89**, issue 24 (1992) 11673-11677.

103. Waksman G., Kominos D., Robertson S.C., Pant N., Baltimore D., Birge R.B., Cowburn D., Hanafusa H., Mayer B.J., Overduin M. and *al. e.* Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* **358** (1992) 625-626.

104. Eck M.J., Atwell S.K., Shoelson S.E. and Harrison S.C. Structure of the regulatory domains of the Src-family tyrosine kinase Lck. *Nature* **368** (1994) 764-769.

105. Tong L., Warren T.C., King J., Batageri R., Rose J. and Jakes S. Crystal structures of the human p56lck SH2 domain in complex with two short phosphotyrosyl peptides at 1.0 Å and 1.8 Å resolution. *Journal of Molecular Biology* **256** (1996) 601-610.

106. Rahuel J., Gay B., Erdmann D., Strauss A., García-Echeverría C., Furet P., Caravatti G., Fretz H., Schoepfer J. and Grütter M. Structural basis for specificity of GRB2-SH2 revealed by a novel ligand binding mode. *Nature Structural Biology* **3**, issue 7 (1996) 586-589.

107. Gay B., Furet P., García-Echeverría C., Rahuel J., Chêne P., Fretz H., Schoepfer J. and Caravatti G. Dual specificity of Src homology 2 domains for phosphotyrosine peptide ligands. *Biochemistry* **36** (1997) 5712-5718.

108. Kimber M.S., Nachman J., Cunningham A.M., Gish G.D., Pawson T. and Pai E.F. Structural basis for specificity switching of the Src SH2 domain. *Molecular Cell* **5**, issue 6 (2000) 1043-1049.

109. García-Echeverría C. Antagonists of the Src Homology 2 (SH2) domains of Grb2, Src, Lck and ZAP-70. *Current Medicinal Chemistry* **8** (2001) 1589-1604.

110. Furet P., Gay B., García-Echeverría C., Rahuel J., Fretz H., Schoepfer J. and Caravatti G. Discovery of 3-aminobenzyloxycarbonyl as an N-terminal group conferring high affinity to the minimal phosphopeptide sequence recognized by the Grb2-SH2 domain. *Journal of Medicinal Chemistry* **40** (1997) 3551-3556.

111. Furet P., Caravatti G., Denholm A.A., Faessler A., Fretz H., García-Echeverría C., Gay B., Irving E., Press N.J., Rahuel J., Schoepfer J. and Walker C.V. Structure-based design and synthesis of phosphinate isosteres of phosphotyrosine for incorporation in Grb2-SH2 domain inhibitors. Part 1. *Bioorganic & Medicinal Chemistry Letters* **10** (2000) 2337-2341.

112. Walker C.V., Caravatti G., Denholm A.A., Egerton J., Faessler A., Furet P., García-Echeverría C., Gay B., Irving E., Jones K., Lambert A., Press N.J. and Woods J. Structure-based design and synthesis of phosphinate isosteres of phosphotyrosine for incorporation in Grb2-SH2 domain inhibitors. Part 2. *Bioorganic & Medicinal Chemistry Letters* **10** (2000) 2343-2346.

113. Schoepfer J., Gay B., Caravatti G., García-Echeverría C., Fretz H., Rahuel J. and Furet P. Structure-based design of peptidomimetic ligands of the Grb2-SH2 domain. *Bioorganic & Medicinal Chemistry Letters* **8** (1998) 2865-2870.

114. García-Echeverría C., Gay B., Rahuel J. and Furet P. Mapping the X(+1) binding site of the Grb2-SH2 domain with α,α-disubstituted cyclic α-amino acids. *Bioorganic & Medicinal Chemistry Letters* **9** (1999) 2915-2920.

115. Schoepfer J., Fretz H., Gay B., Furet P., García-Echeverría C., End N. and Caravatti G. Highly potent inhibitors of the Grb2-SH2 domain. *Bioorganic & Medicinal Chemistry Letters* **9** (1999) 221-226.

116. Caravatti G., Rahuel J., Gay B. and Furet P. Structure-based design of a non-peptidic antagonist of the SH2 domain of Grb2. *Bioorganic & Medicinal Chemistry Letters* **9** (1999) 1973-1978.

117. Tong L., Warren T.C., Lukas S., Schembri-King J., Batageri R., Proudfoot J.R. and Jakes S. Carboxymethyl-phenylalanine as a replacement for phosphotyrosine in SH2 domain binding. *Journal of Biological Chemistry* **273**, issue 32 (1998) 20238-20242.

118. Vu C.B. Recent advances in the design and synthesis of SH2 inhibitors of Src, Grb2 and ZAP-70. *Current Medicinal Chemistry* **7** (2000) 1081-1100.

119. Sawyer T.K., Bohacek R.S., Dalgarno D.C., Eyermann C.J., Kawahata N., Metcalf III C.A., Shakespeare W.C., Sundaramoorthi R., Wang Y. and Yang M.G. Src homology-2 inhibitors: Peptidomimetic and nonpeptide. *Mini Reviews in Medicinal Chemistry* **2**, issue 5 (2002) 475-488.

120. Cussac D., Vidal M., Leprince C., Liu W.-Q., Cornille F., Tiraboschi G., Roques B.P. and Garbay C. A Sos-derived peptidimer blocks the Ras signaling pathway by binding both Grb2 SH3 domains and displays antiproliferative activity. *FASEB Journal* **13** (1999) 31-38.

121. Liu W.-Q., Vidal M., Olszowy C., Million E., Lenoir C., Dhôtel H. and Garbay C. Structure-activity relationships of small phosphopeptides, inhibitors of Grb2 SH2 domain, and their prodrugs. *Journal of Medicinal Chemistry* **47**, issue 5 (2004) 1223-1233.

122. Domchek S.M., Auger K.R., Chartterjee S., Burke Jr. T.R. and Shoelson S.E. Inhibition of SH2 domain/phosphoprotein association by a nonhydrolyzable phosphonopeptide. *Biochemistry* **31** (1992) 9865-9870.

123. Burke Jr. T.R., Smyth M.S., Otaka A., Nomizu M., Roller P.P., Wolf G., Case R. and Shoelson S.E. Nonhydrolyzable phosphotyrosyl mimetics for the preparation of phosphatase-resistant SH2 domain inhibitors. *Biochemistry* **33** (1994) 6490-6494.

124. Yao Z.-J., King C.R., Cao T., Kelley J., Milne G.W.A., Voigt J.H. and Burke Jr. T.R. Potent inhibition of Grb2 SH2 domain binding by non-phosphate-containing ligands. *Journal of Medicinal Chemistry* **42** (1999) 25-35.

125. Burke Jr. T.R. and Lee K. Phosphotyrosyl mimetics in the development of signal transduction inhibitors. *Accounts of Chemical Research* **36**, issue 6 (2003) 426-433.

126. Gao Y., L W., Luo J.H., Guo R., Yang D., Zhang Z.-Y. and Burke Jr. T.R. Examination of novel non-phosphorus-containing phosphotyrosyl mimetics against protein-tyrosine phosphatase-1B and demonstration of differential affinities toward Grb2 SH2 domains. *Bioorganic & Medicinal Chemistry Letters* **10** (2000) 923-927.

127. Gao Y., Voigt J., Xu J.X., Yang D. and Burke Jr. T.R. Macrocyclization in the design of a conformationally constrained Grb2 SH2 domain inhibitor. *Bioorganic & Medicinal Chemistry Letters* **11** (2001) 1889-1892.

128. Gao Y., Wei C.-Q. and Burke Jr. T.R. Olefin metathesis in the design and synthesis of a globally constrained Grb2 SH2 domain inhibitor. *Organic Letters* **3**, issue 11 (2001) 1617-1620.

129. Shi Z.-D., Lee K., Wei C.-Q., Roberts L.R., Worthy K.M., Fisher R.J. and Burke Jr. T.R. Synthesis of a 5-methylindolyl-containing macrocycle that displays ultrapotent Grb2 SH2 domain-binding affinity. *Journal of Medicinal Chemistry* **47** (2004) 788-791.

130. Oishi S., Shi Z.-D., Worthy K.M., Bindu L.K., Fisher R.J. and Burke T.R. Ring-closing metathesis of C-terminal allylglycine residues with an N-terminal ß-vinyl-substituted phosphotyrosyl mimetic as an approach to novel Grb2 SH2 domain-binding macrocycles. *Chemistry & Biochemistry* **6** (2005) 668-674.

131. Oishi S., Karki R.G., Shi Z.-D., Worthy K.M., Bindu L., Chertov O., Esposito D., Frank P., Gillette W.K., Maderia M., Hartley J., Nicklaus M.C., Barchi Jr. J.J., Fisher R.J. and Burke Jr. T.R. Evaluation of macrocyclic Grb2 SH2 domain-binding peptide mimetics prepared by ring-closing metathesis of C-terminal allylglycines with an N-terminal ß-vinyl-substituted phosphotyrosyl mimetic. *Bioorganic & Medicinal Chemistry Letters* **13** (2005) 2431-2438.

132. Shi Z.-D., Wei C.-Q., Lee K., Liu H., Zhang M., Araki T., Roberts L.R., Worthy K.M., Fisher R.J., Neel B.G., Kelley J.A., Yang D. and Burke Jr. T.R. Macrocyclization in the design of non-phosphorus-containing Grb2 SH2 domain-binding ligands. *Journal of Medicinal Chemistry* **47** (2004) 2166-2169.

133. Lee K., Zhang M., Liu H., Yang D. and Burke Jr. T.R. Utilization of a ß-aminophosphotyrosyl mimetic in the design and synthesis of macrocyclic Grb2 SH2 domain-binding peptides. *Journal of Medicinal Chemistry* **46** (2003) 2621-2630.

134. Shi Z.-D., Lee K., Liu H., Zhang M., Roberts L.R., Worthy K.M., Fivash M.J., Fisher R.J., Yang D. and Burke Jr. T.R. A novel macrocyclic tetrapeptide mimetic that exhibits low-picomolar Grb2 SH2 domain-binding affinity. *Biochemical and Biophysical Research Communications* **310** (2003) 378-383.

135. Sayos J., Wu C., Morra M., Wang N., Zhang X., Allen D., van Schaik S., Notarangelo L., Geha R., Roncarolo M.G., Oettgen H., de Vries J.E., Aversa G. and Terhorst C. The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM. *Nature* **395**, issue 6701 (1998) 441-444.

136. Morra M., Simarro-Grande M., Martin M., Chen A.S., Lanyi A., Silander O., Calpe S., Davis J., Pawson T., Eck M.J., Sumegi J., Engel P., Li S.-C. and Terhorst C. Characterization of SH2D1A missense mutations identified in X-linked lymphoproliferative disease patients. *Journal of Biological Chemistry* **276**, issue 39 (2001) 36809-36816.

137. Poy F., Yaffe M.B., Sayos J., Saxena K., Morra M., Sumegi J., Cantley L.C., Terhorst C. and Eck M.J. Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition. *Molecular Cell* **4** (1999) 555-561.

138. Hwang P.M., Li C., Morra M., Lillywhite J., Muhandiram D.R., Gertler F., Terhorst C., Kay L.E., Pawson T., Forman-Kay J.D. and Li S.-C. A 'three-ponged' binding mechanism for the SAP/SH2D1A SH2 domain: structural basis and relevance to the XLP syndrome. *The EMBO Journal* **21**, issue 3 (2002) 314-323.

139. McGuigan C., Devine K.G., O'Connor T.J., Galpin S.A., Jeffries D.J. and Kinchington D. Synthesis and evaluation of some novel phosphoramidate derivatives of 3'-azido-3'-deoxythymidine (AZT) as anti-HIV compounds. *Antiviral Chemistry & Chemotherapy* **1** (1990) 107-113.

140. Burke Jr. T.R., Yao Z.-J., Liu D.-G., Voigt J.H. and Gao Y. Phosphotyrosyl mimetics in the design of peptide-based signal transduction inhibitors. *Biopolymers (Peptide Science)* **60** (2001) 32-44.

141. Burke Jr. T.R., Luo J.H., Yao Z.-J., Gao Y., Zhao H., Milne G.W.A., Guo R., Voigt J.H., King C.R. and Yang D. Monocarboxylic-based phosphotyrosyl mimetics in the design of Grb2 SH2 domain inhibitors. *Bioorganic & Medicinal Chemistry Letters* **9** (1999) 347-352.

142. Long Y.-Q., Voigt J.H., Lung F.-D.T., King C.R. and Roller P.P. Significant compensatory role of position Y-2 conferring high affinity to non-phosphorylated inhibitors of Grb2-SH2 domain. *Bioorganic & Medicinal Chemistry Letters* **9** (1999) 2267-2272.

143. Lung F.-D.T., Long Y.-Q., Roller P.P., King C.R., Varady J., Wu X.W. and Wang S. Functional preference of the constituent amino acid residues in a phage-library-based nonphosphorylated inhibitor of the Grb2-SH2 domain. *Journal of Peptide Research* **57** (2001) 447-454.

144. Lou Y.-G., Lung F.-D.T., Pai M.-T., Tzeng S.-R., Wei S.-Y., Roller P.P. and Cheng J.-W. Solution structure and dynamics of GT1E, a nonphosphorylated cyclic peptide inhibitor for the Grb2 SH2 domain. *Archives of Biochemistry and Biophysics* **372**, issue 2 (1999) 309-314.

145. Long Y.-Q., Yao Z.-J., Voigt J.H., Lung F.-D.T., Luo J.H., Burke Jr. T.R., King C.R., Yang D. and Roller P.P. Structural requirements for Tyr in the consensus sequence Y-E-N of a novel nonphosphorylated inhibitor to the Grb2-SH2 domain. *Biochemical and Biophysical Research Communications* **264**, issue 3 (1999) 902-908.

146. Li P., Zhang M., Peach M.L., Zhang X., Liu H., Nicklaus M., Yang D. and Roller P.P. Structural basis for a non-phosphorus-containing cyclic peptide binding to Grb2-SH2 domain with high affinity. *Biochemical and Biophysical Research Communications* **307** (2003) 1038-1044.

147. Li P., Zhang M., Peach M.L., Liu H., Yang D. and Roller P.P. Concise and enantioselective synthesis of Fmoc-Pmp(Bu^t)$_2$-OH and design of potent Pmp-containing Grb2-SH2 domain antagonists. *Organic Letters* **5** (2003) 3095-3098.

148. Li P., Peach M.L., Zhang M., Liu H., Yang D., Nicklaus M. and Roller P.P. Structure-based design of thioether-bridged cyclic phosphopeptides binding to Grb2-SH2 domain. *Bioorganic & Medicinal Chemistry Letters* **13**, issue 5 (2003) 895-899.

149. Long Y.-Q., Guo R., Luo J.H., Yang D. and Roller P.P. Potentiating effect of distant sites in non-phosphorylated cyclic peptide antagonists of the Grb2-SH2 domain. *Biochemical and Biophysical Research Communications* **310** (2003) 334-340.

150. Song Y.L., Roller P.P. and Long Y.-Q. Development of 1-3-aminotyrosine suitably protected for the synthesis of a novel nonphosphorylated hexapeptide with low-nanomolar Grb2-SH2 domain-binding affinity. *Bioorganic & Medicinal Chemistry Letters* **14** (2004) 3205-3208.

151. Song Y.L., Peach M.L., Roller P.P., Qiu S., Wang S. and Long Y.-Q. Discovery of a novel nonphosphorylated pentapeptide motif displaying high affinity for Grb2-SH2 domain by the utilization of 3'-substituted tyrosine derivatives. *Journal of Medicinal Chemistry* **49**, issue 5 (2006) 1585-1596.

152. Liu W.-Q., Carreaux F., Meudal H., Roques B.P. and Garbay C. Synthesis of constrained 4-(phosphonomethyl)phenylalanine derivatives as hydrolytically stable analogs of O-phosphotyrosine. *Tetrahedron* **52**, issue 12 (1996) 4411-4422.

153. Stankovic C.J., Surendran N., Lunney E.A., Plummer M.S., Para K.S., Shahripour A., Fergus J.H., Marks J.S., Herrera R., Hubbell S.E., Humblet C., Saltiel A.R., Stewart B.H. and Sawyer T.K. The role of 4-phosphonodifuloromethyl- and 4-phosphono-phenylalanine in the selectivity and cellular uptake of SH2 domain ligands. *Bioorganic & Medicinal Chemistry Letters* **7** (1997) 1909-1914.

154. Marseigne I. and Roques B.P. Synthesis of new amino acids mimicking sulfated and phosphorylated tyrosine residues. *Journal of Organic Chemistry* **53** (1988) 3621-3624.

155. Liu W.-Q., Olszowy C., Bischoff L. and Garbay C. Enantioselective synthesis of (2S)-2-(4-phosphonophenylmethyl)-3-aminopropanoic acid suitably protected for peptide synthesis. *Tetrahedron Letters* **43** (2002) 1417-1419.

156. Smyth M.S., Ford Jr. H. and Burke Jr. T.R. A general method for the preparation of benzylic α,α-difluorophosphonic acids; non-hydrolyzable mimetics of phosphotyrosine. *Tetrahedron Letters* **33** (1992) 4137-4140.

157. Burke T.R., Smyth M.S., Nomizu M., Otaka A. and Roller P.P. Preparation of fluoro-4-(phosphonomethyl)-D,L-phenylalanineand hydroxy-4-(phosphonomethyl)-D,L-phenylalanine suitably protected for solid-phase synthesis of peptides containing hydrolytically stable analogues of O-phosphotyrosine. *Journal of Organic Chemistry* **58** (1993) 1336-1340.

158. Yao Z.-J., Gao Y., Voigt J., Ford Jr. H. and Burke Jr. T.R. Synthesis of Fmoc-protected 4-carboxy-difluoromethyl-L-phenylalanine: A phosphotyrosyl mimetic of potential use for signal transduction studies. *Tetrahedron* **55**, issue 10 (1999) 2865-2874.

159. Gao Y., Luo J., Yao Z.-J., Guo R., Zou H., Kelley J., Voigt J.H., Yang D. and Burke Jr. T.R. Inhibition of Grb2 SH2 domain binding by non-phosphate-containing ligands. 2. 4-(2-malonyl)phenylalanine as a potent phosphotyrosyl mimetic. *Journal of Medicinal Chemistry* **43** (2000) 911-920.

160. Kole H.K., Ye B., Akamatsu M., Yan X., Barford D., Roller P.P. and Burke Jr. T.R. Protein-tyrosine phosphatase inhibition by a peptide containing the phosphotyrosyl mimetic, O-malonyl-tyrosine (OMT). *Biochemical and Biophysical Research Communications* **209** (1995) 817-822.

161. Ye B., Akamatsu M., Shoelson S.E., Wolf G., Giorgetti-Peraldi S., Yan X.J., Roller P.P. and Burke T.R. L-O-(2-malonyl)-tyrosine: A new phosphotyrosyl mimetic for the preparation of Src homology 2 domain inhibitory peptides. *Journal of Medicinal Chemistry* **38** (1995) 4270-4275.

162. Burke T.R., Ye B., Akamatsu M., Ford H., Yan X.J., Kole H.K., Wolf G., Shoelson S.E. and Roller P.P. 4'-O-[2-(2-fluoromalonyl)]-L-tyrosine: A phosphotyrosyl mimic for the preparation of signal transduction inhibitory peptides. *Journal of Medicinal Chemistry* **39** (1996) 1021-1027.

163. Burke Jr. T.R., Yao Z.-J., Zhao H., Milne G.W.A., Wu L., Zhang Z.-Y. and Voigt J. Enantioselective synthesis of nonphosphorus-containing phosphotyrosyl mimetics and their use in the preparation of tyrosine phosphatase inhibitory peptides. *Tetrahedron* **54**, issue 34 (1998) 9981-9994.

Principaux résultats

Dynamique moléculaire sur deux complexes de référence

Cette partie résume les principaux points de l'article #1 "Role of water molecules for binding inhibitors in the SH2 domain of Grb2 : a molecular dynamics study". On trouvera le texte de cette publication dans la section suivante.

Situation initiale

De façon classique, une structure cristallographique de résolution suffisante d'un complexe de type protéine / ligand correspond à un point de départ potentiel pour une étude par les techniques de la modélisation moléculaire. Le modélisateur, dès lors que son travail lui permet d'estimer l'affinité d'un ligand donné pour la cible, peut suggérer de nouvelles structures potentiellement intéressantes aux biochimistes, ou au contraire, les dissuader de synthétiser une molécule donnée. Dans un tel contexte, une approche d'optimisation peut être constituée par des simulations par dynamique moléculaire.

Notre point de départ est constitué, au niveau purement structural, de la structure expérimentale d'un complexe Grb2 SH2 / ligand. Ce ligand de référence est noté ligand 1 dans l'article ; on étudie également un dérivé noté ligand-2. Les informations structurales sont complétées par des données de tests biologiques : ligand-1 a une affinité nanomolaire lors de tests *in vitro* sur Grb2 SH2, tandis que les mêmes tests ont pu montrer que l'affinité de ligand-2 s'avérait négligeable en comparaison.

Ligand-1 (haut) et ligand-2 (bas)

Diagramme des interactions du complexe Grb2 SH2 / ligand-1 (structure PDB 1JYQ)

Antérieurement au travail effectué dans le cadre de cette thèse, des simulations par dynamique moléculaires avaient été effectuées sur les complexes de Grb2 SH2 avec ligand-1 et ligand-2. Le mode de liaison résolu expérimentalement pour le complexe de ligand-1 est reproduit correctement. Toutefois, il était également prévu que ligand-2 partage ce même mode de liaison, avec une énergie d'interaction plus importante ; on aurait donc pu prévoir que ligand-2 possède une affinité plus importante pour Grb2 SH2 que ligand-1. Nous observons ici une contradiction évidente entre résultats expérimentaux et prédictions théoriques.

L'objectif de cette thèse étant l'étude théorique du récepteur Grb2 SH2, il s'avéra nécessaire, avant toute chose, de fournir une explication cohérente à cette situation. Bien entendu, les protocoles expérimentaux n'étant pas sujets à caution, nous avons commencé par vérifier et reproduire les simulations effectuées ; aucune erreur dans la modélisation n'a été détectée. On pouvait alors supposer que le système était décrit de façon insuffisante, un paramètre important[*] n'étant pas pris en compte. Nous avons décidé de tester simultanément deux hypothèses : d'une part, le choix du champ de force, d'autre part, les effets de solvant.

Mise en œuvre

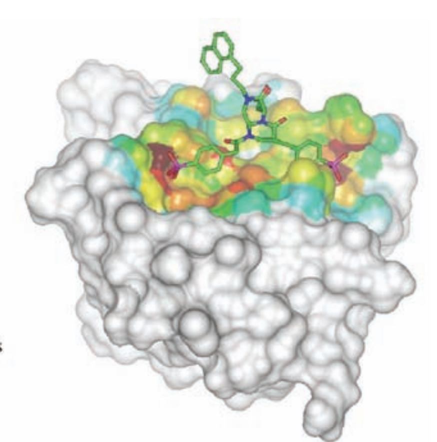
Systèmes modélisés

De la structure expérimentale du complexe Grb2 SH2 / ligand-1 (1JYQ), on conserve un monomère (B), le ligand-1 lié, ainsi que les molécules d'eau situées à moins de 5 Å de la protéine *et* du ligand. Le système du complexe avec ligand-2 est modélisé sur la base du précédent, ligand-2 étant placé par superposition avec ligand-1 ; toutes les molécules d'eau issues de la RX sont supprimées. Enfin, Grb2 SH2 non lié est obtenu en supprimant ligand-2. Les trois types de système sont ensuite immergés dans des boîtes d'eau cubiques de 60 ou 80 Å[†], neutralisés électriquement (contre-ions) et protonés (pH = 7).

Trois champs de force sont sélectionnés : CFF91, AMBER et CHARMM22. Discover est utilisé dans les deux premiers cas, NAMD dans le troisième. Les dynamiques (P = 1 atm, T = 300 K) sont lancées après une phase de minimisation et une phase d'équilibration. Aucune contrainte n'est imposée au système (mis à part la rigidité des liaisons O-H des molécules d'eau).

Nous obtenons sept trajectoires distinctes :

- complexe / ligand 1, CFF91, 1 ns
- complexe / ligand 1, AMBER, 2 ns
- complexe / ligand 1, CHARMM22, 2 ns
- complexe / ligand 2, CFF91, 1 ns
- complexe / ligand 2, AMBER, 1,17 ns
- complexe / ligand 2, CHARMM22, 2 ns
- domaine SH2 non complexé, CHARMM22, 2 ns (pour validation du modèle du récepteur)

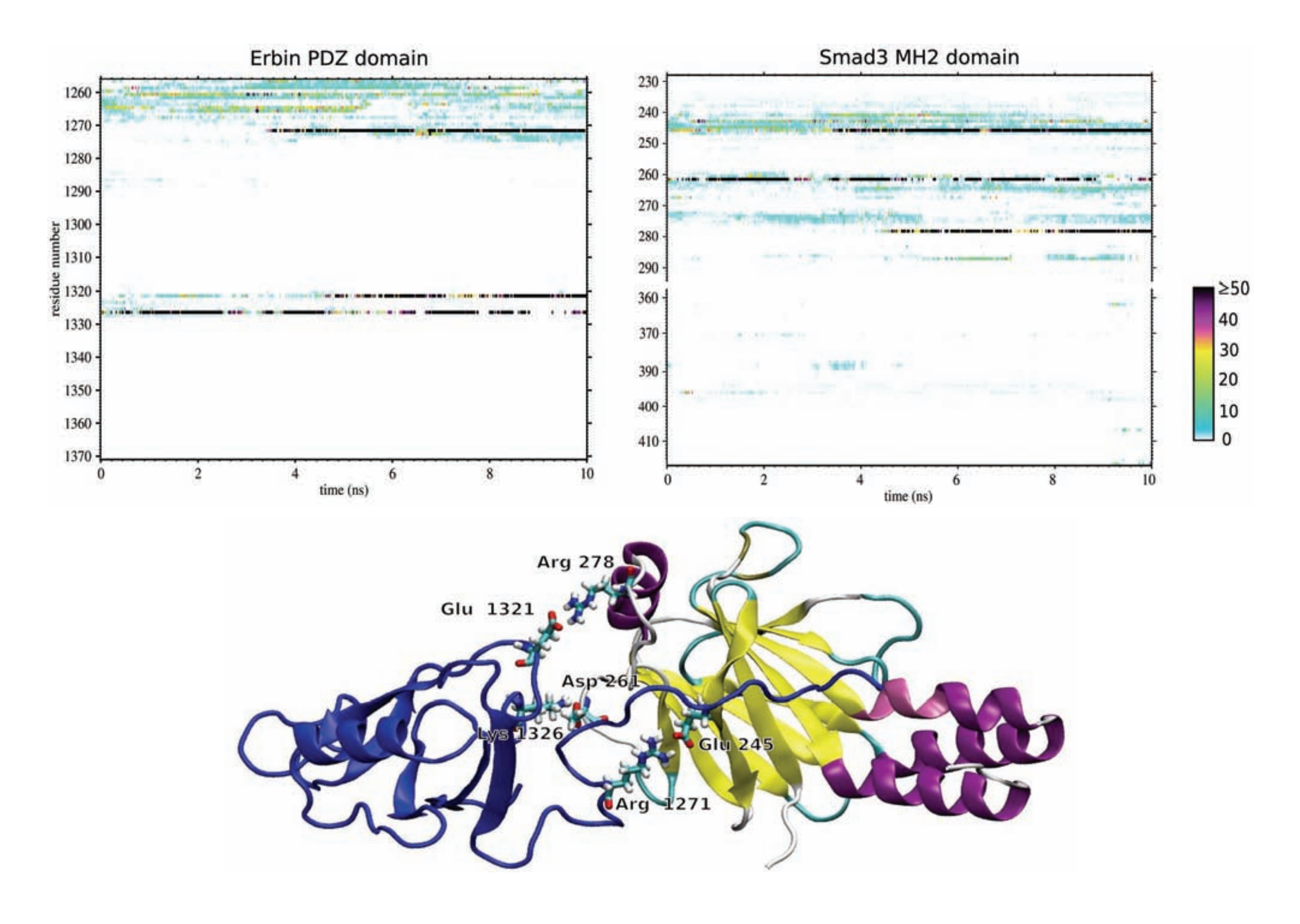


[*] Par exemple, des effets de polarisation de grande amplitude. Ce type de phénomènes marque typiquement les limites des méthodes de la mécanique moléculaire face à celles (en général trop coûteuses pour des systèmes de cette taille) de la mécanique quantique.

[†] Nous avons utilisé Discover et NAMD pour effectuer les dynamiques moléculaires. Il s'avéra que Discover n'est pas capable de gérer un fichier en entrée contenant plus de 10000 résidus. Or chaque molécule d'eau est considérée comme un résidu : il fut impossible de modéliser des boîtes d'eau de 80 Å dans ce cas.

Techniques d'analyse

Outre l'analyse visuelle avec InsightII ou VMD, et les vérifications usuelles (taille de boîte, structure de la chaîne protéique, évolution des distances intermoléculaires intéressantes), deux types de mesures permettant une analyse graphique aisée ont été effectuées sur les trajectoires. Les graphes de RMSD, obtenus avec InsightII, permettent de repérer les changements conformationnels les plus significatifs. Les barres d'énergies, obtenues pour les trajectoires calculées par NAMD, permettent une décomposition des interactions du système, ainsi que la surveillance de leur évolution au cours du temps. Cette technique, qui utilise des programmes et scripts crées spécifiquement, s'avère très utile pour tous types de trajectoires, et a été appliquée avec succès au sein du laboratoire afin de définir aisément une "carte des interactions" pour des simulations de type protéine / protéine (*voir l'exemple ci-dessous*[*]).



Choix du champ de force

L'analyse comparée des trajectoires des deux complexes avec les trois champs de force indique clairement que, sur le plan qualitatif, les résultats d'un champ de force à l'autre sont similaires. De façon satisfaisante, les trois simulations du complexe ligand-1 confirment la stabilité de celui-ci, tandis que les complexes de ligand-2 s'avèrent instables. Les *cluster graphs* des 6 trajectoires des complexes indiquent toutefois que des différences mineures peuvent être observées. On peut ainsi noter, pour les trajectoires du complexe ligand-1 avec les champs de force CFF91 et AMBER, une légère réorganisation structurale postérieure à la phase d'équilibration, qui n'apparaît pas dans la dynamique avec CHARMM22. La trajectoire obtenue avec ce dernier champ de force reste donc la plus stable par rapport à la structure expérimentale de départ, même si on peut considérer que la stabilité avec les deux autres champs de force s'avère tout à fait satisfaisante.

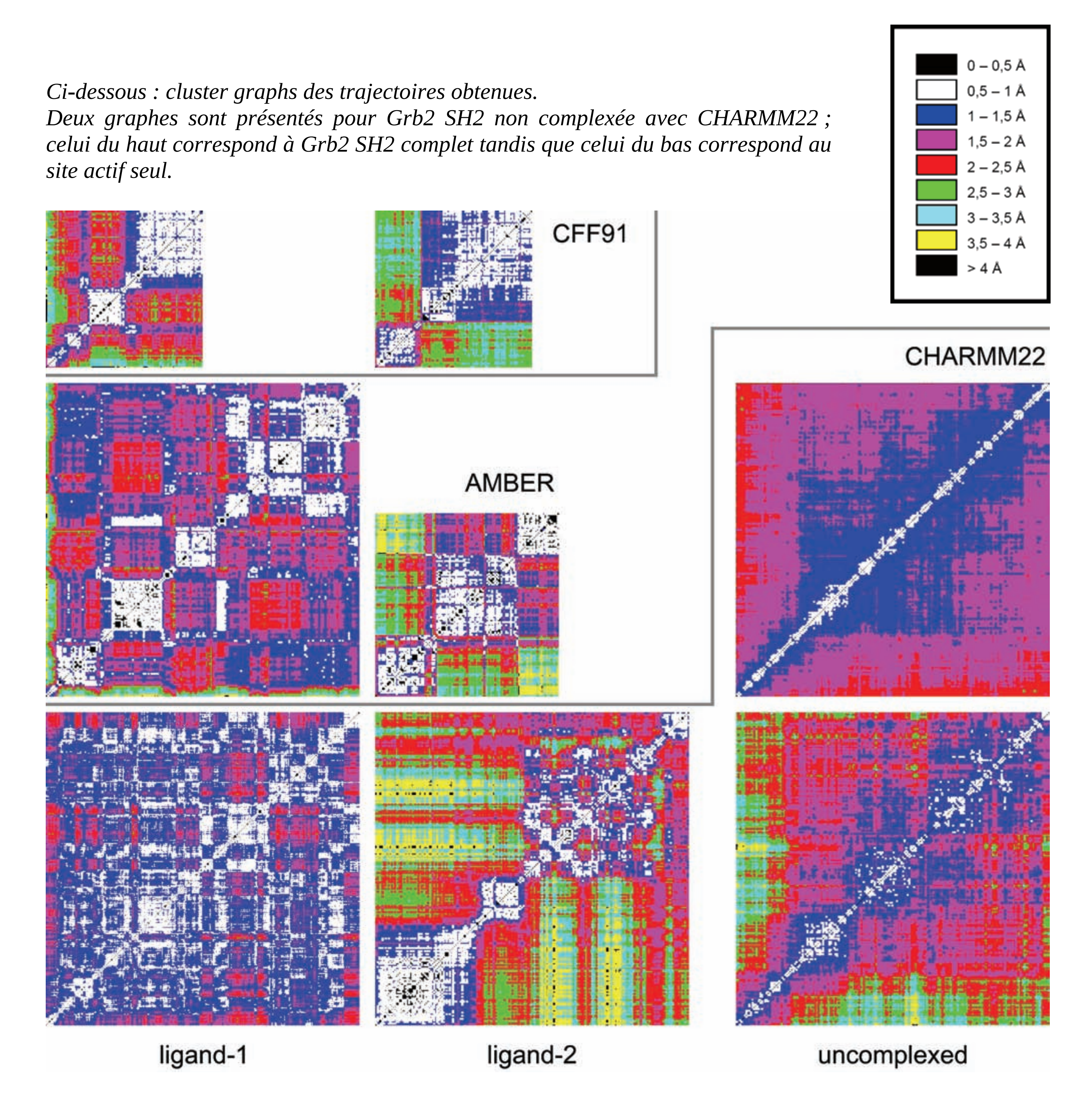
[*] Merci à Matthieu Chavent pour ces images.

Il n'a pas été jugé bon de procéder dans l'article à une analyse structurale détaillée de toutes les trajectoires. Nous avons choisi arbitrairement de nous focaliser sur les dynamiques obtenues avec CHARMM22 / NAMD, d'une part pour des raisons de clarté, et d'autre part car l'analyse des résultats de NAMD est facilitée (barres d'énergie).

Ainsi (1) le protocole de simulation mis en place reproduit les résultats expérimentaux, et (2) l'hypothèse du choix du champ de force comme cause de la non représentativité des précédentes simulations est invalidée (du moins pour les trois champs de force employés).

Validation du protocole

L'analyse de la trajectoire de la protéine Grb2 non complexée (avec CHARMM22) permet de vérifier la pertinence du protocole employé. La géométrie de la protéine évolue de façon cohérente : la conformation des résidus du site actif évolue afin de maximiser les interactions intramoléculaires, ce qui compense le retrait des interactions protéine / ligand. Aucun changement notable n'est observé en dehors du site actif.



Ci-dessous : cluster graphs des trajectoires obtenues.
Deux graphes sont présentés pour Grb2 SH2 non complexée avec CHARMM22 ; celui du haut correspond à Grb2 SH2 complet tandis que celui du bas correspond au site actif seul.

Nouvelles connaissances concernant l’interaction de ligands sur le récepteur Grb2 SH2

Effets du solvant

La modélisation explicite du solvant constituant le seul ajout méthodologique au protocole des simulations, le fait que celles-ci soient désormais en conformité avec les observations expérimentales accrédite l’hypothèse d’un rôle actif de l’eau au niveau de la liaison de ligands sur le domaine SH2 de Grb2.

Dans le cas de la dynamique du complexe Grb2 SH2 / ligand-1, les molécules d’eau issues de la structure cristallographique ne s’avèrent pas liées fortement à Grb2 ou au ligand, tandis que le positionnement de certaines d’entre elles dans leur voisinage est conservé. Cela semble traduire la stabilité du complexe plutôt que l’expression d’un rôle structural notable de ces molécules d’eau, étant donné que le mode de liaison du ligand s’avère identique à celui prédit antérieurement *in vacuo* et conforme à la structure RX.

L’analyse de la trajectoire de la dynamique du complexe ligand-2 révèle quant à elle que :

- Le premier groupe phosphaté a un mode de liaison alternatif qui n’est pas entièrement localisé dans la cavité-1 (*voir article et figure page 33*). En effet, si le résidu R_{86} est bien ciblé, l’interaction avec R_{67} est considérablement affaiblie au profit de K_{109}, constituante de la cavité-2. Cette dernière interaction s’annule à t ~ 1,8 ns, sans que le mode de liaison de ligand-1 impliquant R_{67} soit réobtenu (*voir figure 10 de l’article*).

- La liaison du second groupe phosphaté dans cavité-3 se fait exclusivement avec le résidu R_{142}. Cela permet à t ~ 1 ns un changement conformationnel défavorable (*bien visible sur la figure 5*) au cours duquel R_{142} se ”déplie” vers le solvant en ”entraînant” ligand-2 (*figure 6*). Ce mouvement concerté optimise cette interaction au détriment de celles impliquant S_{141} et N_{143} (*figure 10*) : au final, environ 100 kcal.mol^{-1} d’interactions avec ligand-2 sont transférés de Grb2 SH2 vers le solvant (*figures 7 et 9*).

Propositions au niveau du design de nouveaux ligands

Les interactions de l’atome d’oxygène non terminal du groupe phosphate des deux phosphotyrosine du ligand-1 s’avèrent non négligeables ; ces deux atomes ne sont pas présents dans ligand-2. En conséquence, bien que le choix de deux groupes phosphone pour ligand-2 à la place des phosphates soit parfaitement justifié sur le plan pharmaceutique[*], on peut émettre l’hypothèse que cette différence structurale soit directement à l’origine aussi bien du mode de liaison alternatif dans cavité-1 que du changement conformationnel défavorable au niveau de cavité-3.

De plus, on a pu observer que les groupes en position -1 ne conféraient aucune activité sur le récepteur Grb2 SH2. La substitution ou la suppression de ces groupes est donc proposée. Ces deux modifications s’avérant à priori simples chimiquement, la synthèse et le test de l’activité biologique d’un dérivé de ligand-2 sans le groupe naph et présentant deux résidus phosphotyrosine non modifiés (*voir ci-contre*) s’avérerait sans doute plus probant que la conduite d’une simulation de dynamique moléculaire supplémentaire.

Proposition de dérivé pour ligand-2

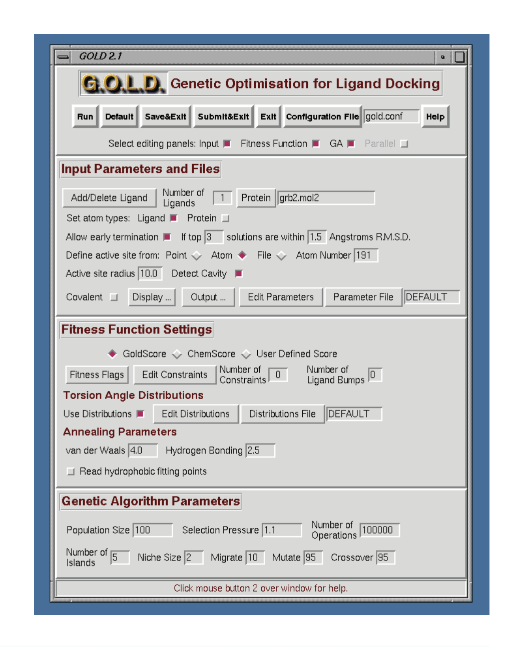
---

[*] Un phosphone étant beaucoup plus stable face aux réactions de déphosphorylation qui peuvent survenir dans l’organisme qu’un phosphate, ligand-2 a un caractère *drug-like* nettement plus important que ligand-1.

Docking flexible sur des bases de molécules

Situation initiale

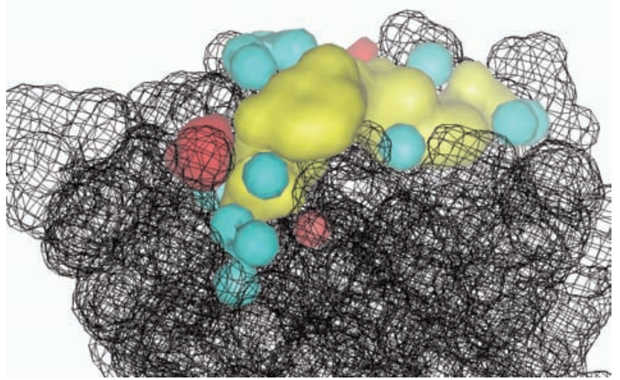
Prise en charge du solvant



Les résultats de dynamique moléculaire allant dans le sens d'un impact potentiellement important du solvant lors de la liaison d'un ligand sur Grb2 SH2, il s'avéra nécessaire de tenir compte de cet aspect au niveau de la mise en place d'un protocole de docking.

La plupart des programmes de docking ne prévoient pas une telle possibilité, et dans le cas de GOLD (version 2), que nous avons utilisé, les interactions caractérisant la désolvatation du récepteur à la liaison du ligand sont favorisées à posteriori.[*] Le fait qu'un ligand doive impérativement "chasser" les molécules d'eau du site actif est traditionnellement considéré comme une condition nécessaire, même si certains systèmes (en particulier HIV protéase) sont connus pour présenter au moins une molécule d'eau structurellement importante.

Afin de tester les effets de l'inclusion du solvant dans le modèle de docking, nous avons retenu les molécules d'eau présentes dans la structure RX de référence du complexe Grb2 SH2 / ligand-1, en modifiant le fichier PDB afin de faire en sorte que GOLD les considère comme des résidus liés de façon covalente à Grb2 (*ci-contre, la structure correspondante ; ligand-1 est en jaune et les molécules d'eau en bleu, tandis que les zones en rouge indiquent des volumes vacants à l'interface*).



Test et validation des paramètres

Afin de tester la fiabilité du protocole de docking, il convient de l'éprouver en vérifiant que les structures de complexes connues expérimentalement sont reproduites.[†] Il s'avéra, dans le cas de GOLD, que seul le modèle du récepteur de Grb2 SH2 défini explicitement, incluant les molécules d'eau cristallographiques, associé aux paramètres standard et un nombre d'échantillonnage de 50, permettait de reproduire les conformations des complexes expérimentaux (1BMB, 1JYR, 1JYW, 1ZFP).

---

[*] À noter que dans sa version 3, GOLD propose un algorithme qui, sur la base d'une liste de positions possibles (typiquement connues par cristallographie RX), détermine la liste des molécules d'eau sur le site actif qui seraient probablement concervées pour une conformation donnée d'un ligand, et en tient compte au niveau de la valeur du score. Il s'agit d'une évolution particulièrement bienvenue. Cette fonctionnalité n'était malheureusement pas disponible lorsque les simulations présentées ici ont été conduites.

[†] Il s'agit d'une nécessité absolue dans la mesure où les tests de validation basés sur cette démarche ne donnent, pour les meilleurs programmes, que des taux de réussite de l'ordre de 60-70%. De plus, l'efficacité d'un programme de docking donné dépend de la nature du système qu'on lui applique, il fallait donc vérifier que GOLD pouvait bien s'appliquer à Grb2 SH2, et dans quelles conditions.

Limitations

L'utilisation d'un programme de docking impose fondamentalement des limitations plus importantes que celles des techniques de la dynamique moléculaire. Principalement, les effets dynamiques sont par nature inaccessibles, une fonction de score empirique et approximative est employée au lieu d'un calcul énergétique faisant intervenir un champ de force, et enfin l'espace de recherche conformationnel est restreint, en particulier au niveau du récepteur. Le programme GOLD a été choisi, car il tient compte de la flexibilité des ligands, dispose d'un algorithme de recherche évolutionnel performant, et propose un rapport intéressant entre fiabilité et vitesse de traitement.[*]

La prise en charge du solvant proposée ici se limite au positionnement de molécules d'eau interfaciales. En aucun cas un tel modèle ne pourra prédire des phénomènes de déstabilisation du *bulk* tels que ceux observés lors des dynamiques de ligand-2 avec Grb2 SH2. Ainsi il est à prévoir, dans le meilleur des cas, que de telles situations donnent lieu au même type de faux positifs que ceux qui seraient probablement obtenus avec une dynamique moléculaire *in vacuo*.

De plus, le fait que l'inclusion des molécules d'eau cristallographiques améliore la précision du docking peut paraître surprenant étant donné que celles-ci n'apparaissaient pas jouer un rôle important dans les trajectoire des dynamiques. On peut donc émettre l'hypothèse que cela corresponde non pas à une amélioration de la pertinence du modèle, mais plutôt à une contrainte structurale limitant l'espace de recherche. En effet, en positionnant ces molécules d'eau, l'étendue du site actif est restreinte, ce qui peut grandement faciliter la tâche du programme de docking. Il faut alors garder à l'esprit que cela pourrait aussi artificiellement défavoriser le docking de molécules possédant un mode de liaison préférentiel différent de celui de ligand-1.

Cette crainte peut toutefois être relativisée. D'une part, l'interaction de ligands sur les domaines SH2 est spécifique et en ce qui concerne Grb2, comme nous l'avons vu précédemment, un seul mode de liaison en *ß-turn* est connu. On peut raisonnablement supposer, dans de telles conditions, qu'aucun autre mode ne soit favorable – même si l'analyse des dynamiques du complexe de ligand-2 semble indiquer que ce point mériterait de plus amples investigations. D'autre part, il est évident qu'afin de pouvoir prédire par docking la liaison d'un ligand dans un mode radicalement différent, il faudrait déjà disposer de la conformation du récepteur correspondante, sachant que les effets d'*induced fit* délimitent une limitation importante de la quasi-totalité des programmes de docking disponibles (dont GOLD).

Screening virtuel sur la cible Grb2 SH2

Molécules sélectionnées comme candidates

Nous avons eu recours à différentes bases moléculaires existantes, provenant de différentes sources (méta-base ZINC, fournisseurs de produits chimiques, collaborateurs) afin de sélectionner les molécules à tester par docking sur Grb2 SH2. Les ligands de Grb2 SH2 connus, des dérivés de ceux-ci, ainsi qu'un certain nombre de molécules construites manuellement, ont été regroupés au sein d'une petite base ciblée de référence.

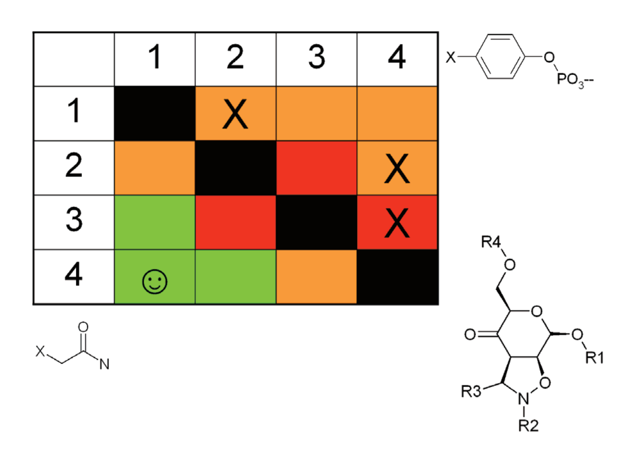
Au total, environ 30000 molécules ont été préparées (si nécessaire, correction d'erreurs, suppression de parasites et conversion en 3D, puis protonation) avant d'être dockées sur Grb2 SH2.

---

[*] Au moment de déterminer le protocole de docking, GOLD semblait être le programme potentiellement le plus précis ; nous nous sommes contentés de vérifier qu'il l'était suffisamment pour notre cible en utilisant les structures expérimentales connues. Les techniques de docking sont en constante évolution et de nouveaux programmes apparaissent régulièrement. Même si GOLD a connu des améliorations notables (en particulier au niveau de la prise en charge du solvant), son intérêt est actuellement nettement moins évident, face aux versions récentes de programmes comme ICM ou Glide. On peut donc dire par rapport aux calculs de docking que s'ils étaient à refaire, il faudrait sans doute avant toute chose prendre plus de temps pour estimer, dans la mesure du possible, l'efficacité des différentes solutions disponibles. L'article 3 décrit une partie des problématiques associées à une telle situation.

Résultats

Une des premières simulations de docking, après validation du protocole, consista, à partir d'un squelette moléculaire proposé par des expérimentateurs, possédant quatre zones de substitution pour deux substituants possibles, à déterminer la position optimale de ces derniers pour Grb2 SH2. Il s'avéra clairement que les positions les plus favorables (*en vert ci-contre*) correspondaient précisément aux synthèses les plus délicates (*ci-contre, les croix indiquent les substitutions envisageables pour synthèse*). Le squelette en question a finalement été abandonné.



Parmi les ~30000 molécules dockées par la suite, de façon prévisible les molécules construites manuellement à partir de ligand-1 et d'autres ligands actifs connus produisent les meilleurs résultats. Plusieurs modifications suggérées par l'analyse des dynamiques de ligand-1 donnent lieu à des scores améliorés. De façon cohérente avec les connaissances préalables sur l'inhibition de Grb2 SH2, la séquence de référence pYVNV interagit moins fortement que ligand-1, tandis que pYpYN dont ligand-1 est issu est encore moins actif (orientation du résidu pY_{+1} non optimisée).

Différentes molécules disposant d'une charge électrique inférieure ou égale à celle de la référence et ne possédant pas de groupe phosphate instable sont mises en évidence. Ces structures pourraient s'avérer utiles en tant que bases pour la mise au point d'inhibiteurs de Grb2 SH2 plus intéressants en tant que médicaments potentiels. La structure des complexes correspondants prédite par docking peut constituer la base d'optimisations par modélisation moléculaire.

Molécules intéressantes issues des calculs de docking flexible, classées suivant leur score GOLD vis-à-vis de Grb2 SH2 (décroissant de haut en bas). De gauche à droite, molécules neutres, chargées 1^-, 2^- et 4^-. Les molécules encadrées en bleu sont issues de la base ciblée "eDAM" (on remarque ligand-1 situé en seconde position en haut à droite), celles encadrées en jaune de la base "Yao" (pharmacopée chinoise) ; enfin la molécule encarée en vert provient de la base ZINC.

Validation

En tant que tels, ces résultats de docking n'ont qu'une faible signification. En l'absence de tests expérimentaux des structures suggérées, le protocole de docking doit être validé sur la base de résultats existants, avant d'envisager le test d'un plus grand nombre de molécules.

Nous disposons pour ce faire de résultats détaillés d'expériences de screening réalisées sur différents domaines SH2, dont Grb2. Les domaines SH2 reconnaissant tous une ou plusieurs séquences du type pYXXX, le screening des 8000 combinaisons possibles a été effectué, et les "préférences" précises à chaque position +1, +2 et +3 sont connues.*

Le protocole de validation est basé sur 4 structures RX de précision suffisante de récepteurs SH2 complexés. Les 8000 peptides pYXXX sont dockés sur chacun de ces récepteurs, sous deux formes, avec et sans conservation des molécules d'eau cristallographiques†, soit un total de 64000 dockings.

Code PDB	Protéine (séquences peptidiques préférées)	Ligand (mode de liaison)
1IS0	Src (pYEEI)	pYEEI (linéaire)
1P13		pYANF (ß-turn)
1LKK	P56-Lck (pYEEI)	pYEEI (linéaire)
1JYQ	Grb2 (pYVNV, pYENW)	pYpY*N (ß-turn)

Les résultats obtenus ne permettent à l'heure actuelle aucune validation : il est difficile de distinguer la moindre spécificité parmi les 8000 combinaisons ; la distribution des scores est étroite, et en particulier, en ce qui concerne Grb2, la spécificité du résidu N_{+2} n'est pas mise en évidence.

Situation présente

Les résultats de docking peuvent difficilement donner lieu à une publication étant donné qu'aucune validation n'est donnée quant à la capacité de GOLD à classer correctement les molécules entrantes suivant leur affinité pour Grb2 SH2. Toutefois, nous conservons les résultats du protocole de validation. On peut émettre l'hypothèse que l'échec actuel de la validation soit dû à la précision insuffisante de la fonction de score de GOLD, point faible souvent évoqué pour ce programme. Les structures obtenues semblent correctes : les peptides pYXXX sont correctement orientés et le groupe $pTyr_0$ est bien positionné dans sa cavité spécifique. Si à l'avenir un protocole automatisé de post-traitement est mis en place (par exemple, par minimisation en utilisant un champ de force), il sera utile l'utiliser afin de tenter de valider le docking, tout en précisant éventuellement à quels niveaux le protocole actuel pourrait être amélioré.

De plus, pour des molécules ponctuelles, la condition à la conduite de simulations de mécanique moléculaire ultérieures plus précises ne requiert qu'une structure satisfaisante. Dans un contexte d'optimisation et non de screening, l'imprécision possible d'un score de docking n'est pas handicapante. On peut donc envisager, même si la validation décrite précédemment n'est pas encore menée à bien, de tester par dynamique moléculaire les structures les plus intéressantes présentées à la page précédente. Cela présente le risque de gaspiller une quantité importante de temps de calcul sur des faux négatifs, mais une telle situation n'est pas inhabituelle dans le cadre général du *drug design.*

---

* On se réfère ici aux publications suivantes:
Songyang *et al. Cell* **72** (1993) 767-778.
Songyang *et al. Mol.Cell. Biol.* **14** (1994) 2777-2785.
Gram *et al. Eur. J. Biochem.* **246** (1997) 633-637.
Kessels *et al. PNAS* **99** (2002) 8524-8529.
Vetter *et al. Curr. Prot. Pept. Sci.* **3** (2002) 365-397.

† Si la conservation de ces molécules d'eau semble être utile pour le docking sur Grb2 SH2, nous n'étions pas en mesure d'en déduire si ce serait ou non le cas pour les trois autres récepteurs SH2.

Screening virtuel : mise en place d'outils innovants

Principales problématiques liées au screening virtuel

Cette partie se place dans le contexte détaillé par l'article #3 : "Should structure-based virtual screening techniques be used more extensively in modern drug discovery ?".

Généralement, lorsque le screening virtuel a pour simple visée de servir de complément à un screening expérimental, son usage s'avère assez simple et une telle approche est devenue assez répandue, car elle permet souvent d'obtenir à moindre coût une quantité d'informations *suffisante* lors des premières phases d'une recherche de médicaments (la découverte de *hits*). Lors des phases plus avancées, les techniques de screening virtuel peuvent apporter un enrichissement des connaissances. Par exemple, sur la base d'un *hit*, il peut être procédé à la génération d'une base de dérivés par chimie combinatoire virtuelle. Cette base peut ensuite être screenée, par exemple en suivant les protocoles classiques de docking déjà employés afin de déterminer les *hits*. Une analyse statistique des résultats peut enfin apporter des précisions permettant d'optimiser les structures intéressantes, s'intégrant ainsi parfaitement aux différentes méthodes de la phase de *lead optimization*.

Ainsi, les techniques de screening virtuel sont considérées comme efficaces lorsqu'elles accompagnent d'autres approches ; elles n'ont alors pas un rôle central dans un protocole de recherche de médicaments. Pour que cela puisse être le cas, il faudrait qu'elles puissent combiner efficacité et fiabilité. Si des méthodes théoriques relativement fiables existent, elles sont délicates à mettre en œuvre. Par exemple, les simulations de dynamique moléculaire nécessitent des informations structurales précises souvent délicates à obtenir pour des assemblages biologiques de taille importante, et exigent de larges moyens de calculs qui les excluent du champ d'application du screening virtuel. Ne sont principalement accessibles que des programmes de docking ou des approches empiriques et statistiques, méthodes fondamentalement approchées. On considère donc souvent, à raison, que telles qu'elles les méthodes de screening virtuel ne sauraient se substituer totalement aux approches classiques de la recherche pharmaceutique.

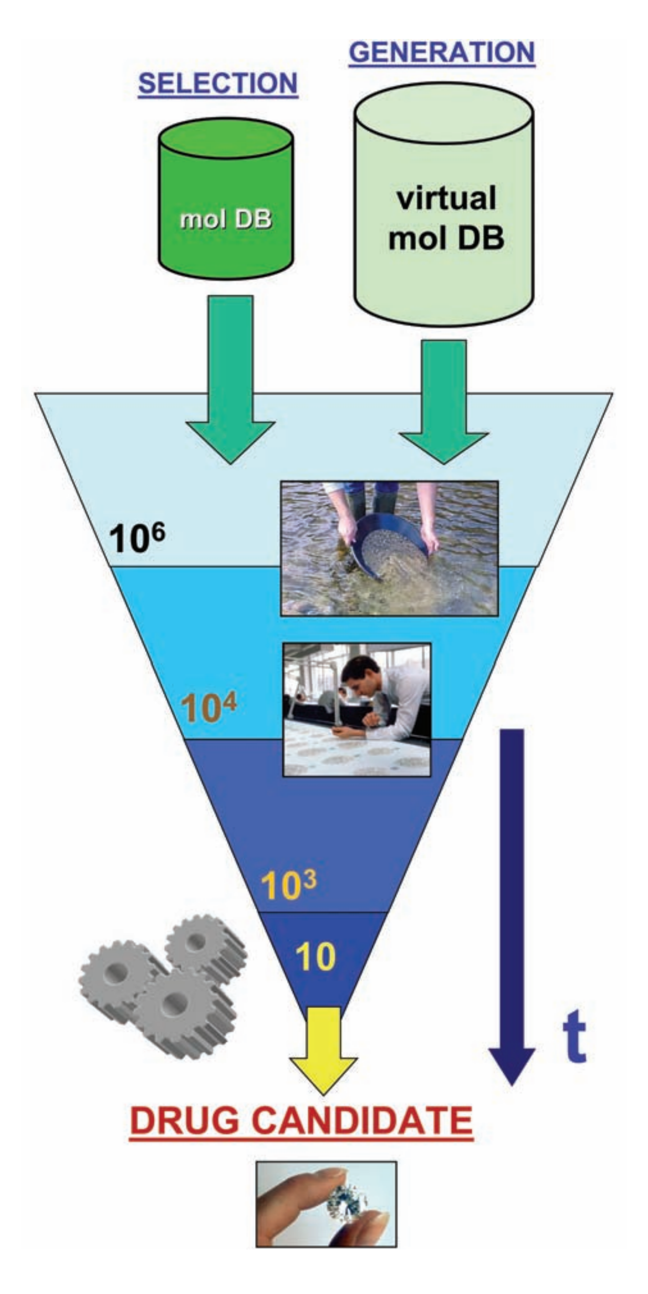
Il existe à l'heure actuelle un très large choix de programmes de docking. À chacune de ces approches correspond une modélisation spécifique au niveau de la flexibilité des structures, de l'algorithme d'exploration de l'espace conformationnel, ainsi que de la fonction de score utilisée pour définir celui-ci. De plus, la possible prise en compte du solvant, aspect qui, comme nous l'avons vu, peut s'avérer déterminant dans certains cas, est souvent insuffisante au-delà des limitations intrinsèques du docking. De façon remarquable, une analyse détaillée du fonctionnement interne des principaux programmes de docking révèle que chacun possède des points forts et des points faibles caractéristiques. Cette hétérogénéité est à double tranchant : si elle peut conférer une certaine liberté à quiconque possède suffisamment d'expertise, elle induit une complexité qui limite l'usage du screening virtuel et est susceptible d'être à l'origine d'erreurs parfois difficilement décelables.

Les recherches théoriques que nous voulons conduire s'orientent à la fois vers la mise en place d'infrastructures combinant les point forts des différentes méthodes disponibles (plutôt que de chercher systématiquement à en créer de nouvelles), et vers l'emploi de protocoles de validation plus stricts (ne se limitant pas à la seule reproduction d'un petit nombre de complexes expérimentaux de référence). Un objectif à long terme consiste à garantir qu'un screening virtuel puisse identifier *tous* les *hits* d'une base de molécules pour une cible, et non plus au moins un avec une bonne probabilité, comme actuellement.

Le projet VSM-G

Cette partie décrit brièvement le projet VSM-G. Une description plus détaillée de VSM-G, ainsi que son protocole de validation sont présents dans l'article #2 "Multiple-step virtual screening using VSM-G : overview and validation of fast geometrical matching enrichment".

Présentation



Comme nous venons de le voir, les techniques de screening virtuel présentent une hétérogénéité très importante qui s'exprime au niveau de leurs conditions d'utilisation, possibilités d'application et pertinence. Cette caractéristique fondamentale peut s'avérer être bénéfique dès lors qu'elle est explicitement prise en compte, afin qu'à un besoin spécifique l'approche la plus appropriée soit sélectionnée.

La mise en place d'un protocole de screening virtuel fiable et suffisamment efficace pour gérer des larges bases de molécules (> 10^6), et pouvant être employé de façon routinière au-delà d'un simple accompagnement des techniques expérimentales, est un défi qu'il est tentant de relever si l'on considère l'avantage que peut conférer les progrès technologiques et conceptuels de l'informatique. Or aucune technique de screening virtuel ne réunit seule les critères de fiabilité et de rapidité de traitement ; on doit donc envisager l'utilisation combinée de plusieurs techniques. Nous avons vu qu'une méthode de docking est d'autant plus approchée que le nombre de dimensions du système est faible. Dans ces conditions, les méthodes les plus rapides ne sont pas aptes à établir un classement précis des molécules testées mais sont à priori capables d'en exclure une partie suffisante. En conséquence, l'utilisation d'une séquence de techniques, de la plus rapide à la plus précise, chacune constituant non pas un évaluateur mais un filtre moléculaire, est envisageable. On obtient ce qu'on peut assimiler à un "entonnoir de screening" (*voir schéma de principe ci-contre*).

À chaque étape de l'entonnoir un nombre significatif de molécules est éliminé du screening virtuel. Les molécules exclues en premier sont celles dont on peut mesurer le manque d'intérêt pour la cible sur la base des modèles les plus simples. Globalement, on cherche à s'assurer que lorsqu'une molécule est exclue à une étape donnée, on a pour ce faire dépensé le minimum de temps de calcul théorique. Le temps de calcul total est ainsi optimisé dès lors que les différentes techniques et paramètres sont bien choisis.

VSM-G (Virtual Screening Manager for computational Grids) constitue fondamentalement une infrastructure logicielle permettant d'implémenter aisément un tel protocole de screening virtuel, et ce de façon totalement automatisée.

Dans ce contexte, l'hétérogénéité des programmes disponibles induit une grande complexité, et peut s'avérer handicapante. Un des principaux objectifs fixé à VSM-G est donc de faciliter autant que possible la conception par l'utilisateur d'un protocole multi-étapes efficace. Pour ce faire, le fonctionnement du *funnel* est automatisé et transparent[*], tandis que l'interface graphique est conçue dans une optique de simplicité[†].

Développements en cours et objectifs

La conduite d'une simulation avec VSM-G s'organise globalement en trois étapes : la préparation de la base des ligands, la préparation de la cible (qui peut être représentée par un nombre quelconque de structures), et enfin le funnel. À l'heure actuelle, cette dernière partie est limitée à trois étapes de docking[‡], et les deux autres sont de simples collections de modules. Le développement de VSM-G n'en est qu'à ses débuts. On trouvera toutefois dans l'article un exemple d'application suffisamment probant pour justifier les investissements consentis jusqu'à présent, tout en soulignant la nécessité de poursuivre cet effort.

Les objectifs directs se situent au niveau (1) de l'utilisation des méthodes de calcul massivement distribuées (grilles de calcul, réseaux hétérogènes), (2) de l'intégration de l'ensemble du spectre des méthodes de la chemoinformatique et de la bioinformatique (méthodes statistiques et empiriques, fouille de données…), et (3) vers une conception plus modulaire (les modules du funnel sont pour le moment "codés en dur", alors que leur implémentation devrait se faire au niveau de l'interface utilisateur).

Perspectives à long terme : vers un screening virtuel "intelligent"

Chaque technique de screening virtuel est susceptible de produire un certain taux de faux positifs et de faux négatifs. Si un faux positif constitue une simple perte de temps car il sera à priori détecté ultérieurement, en particulier dans le contexte du funnel implémenté par VSM-G, l'exclusion à tort d'une molécule est bien plus problématique, car et à priori indécelable. Une évolution conceptuelle de VSM-G devra donc consister à substituer au mécanisme de filtrages successifs par les modules du funnel un système de prioritisation[§], complété par des algorithmes de détection de faux négatifs. Une telle détection pourrait impliquer le "passage forcé" d'une petite proportion de molécules rejetées afin de valider la fiabilité d'un filtre par les prédictions de filtres ultérieurs. La mise en place d'un tel procédé nécessite une analyse en temps réel des résultats, l'implémentation de méthodes de *clustering* de bases moléculaires[**], et peut déboucher sur une optimisation à la volée des paramètres du funnel.[††] De tels mécanismes d'adaptation et d'apprentissage s'intègrent naturellement à l'objectif d'étendre VSM-G à une plateforme d'acquisition de connaissances.

---

[*] Cela nécessite une expertise des différents programmes employés et l'intégration de codes de manipulation de données, plus particulièrement afin d'assurer la conversion entre les formats de fichiers utilisés.

[†] Idéalement il faudrait que VSM-G puisse être utilisé aisément par des scientifiques ayant peu ou pas d'expérience en modélisation moléculaire et en chemoinformatique, le rendant accesible en dehors du cercle des théoriciens. Nous n'en sommes pas encore là !

[‡] On trouvera dans l'article plus de détails concernant ces trois modules, ainsi que sur les autres composantes de VSM-G.

[§] On utilise alors le funnel tant qu'il reste du temps de calcul disponible ; et si ce temps est infini, toutes les molécules de départ passeront l'ensemble des modules. Un tel protocole peut être employé sur de petites bases pour tester l'efficacité de différents modules, et ainsi estimer plus précisément les paramètres de filtrage optimaux.

[**] On peut supposer que la probabilité de faux négatifs d'une méthode donnée et corrélée à des paramètres physico-chimiques précis. Il convient donc de maximiser la diversité des molécules testées pour la détection d'erreurs, tandis que lorqu'un faux négatif est mis en évidence, il faudra tester en priorité les molécules rejetées les plus similaires.

[††] Par exemple, on peut augmenter le taux de rejet d'un module par petits paliers tant qu'aucun faux négatif n'est détecté. Dans le cas contraire, il faut le diminuer, et éventuellement supprimer le module du funnel s'il s'avère que le rapport entre capacité de filtrage et vitesse de traitement devient insuffisant face à un autre module.

Conclusions

Dans le cadre de cette thèse, j'ai été amené à conduire aussi bien des simulations très précises de dynamique moléculaire que des calculs de docking flexible à l'échelle du screening virtuel. J'ai enfin participé à la mise en place de prototypes de protocoles à plus haut débit. Ce travail couvre ainsi les deux étapes de la recherche pharmaceutique auxquelles les simulations numériques de systèmes biomoléculaires peuvent contribuer : la phase de découverte de *hits*, suivie de la phase d'optimisation de *leads*. Certains des obstacles que l'on peut rencontrer en utilisant ces méthodes ont été mis en évidence.

Lors de l'étape d'optimisation de molécules de référence, l'usage de méthodes poussées telles que la dynamique moléculaire est un choix naturel pour le modélisateur. On s'attend généralement par ce biais à décrire de façon précise l'interaction entre deux composés d'intérêt biologique, et la question de savoir si cette interaction est suffisante est souvent résolue à ce stade de la recherche. Cette étude a pourtant débuté par une contradiction à ce niveau précis. Il a été possible de surmonter ce problème en mettant en évidence les effets de solvant, qui, pour la cible Grb2 SH2 et un ligand en particulier, s'avèrent cruciaux. La prise en compte de ce paramètre ouvre la voie à des études ultérieures plus fiables d'un tel système , d'autant plus que le protocole de dynamique moléculaire mis en place pour Grb2 SH2, que l'on peut considérer comme satisfaisant en l'état actuel de nos connaissances, est aisément réutilisable.

Les études de docking ne prennent en compte le solvant que de façon très superficielle. Cet aspect n'est pas nécessairement redhibitoire étant donné que la technique utilisée est par nature approchée à bien d'autres niveaux. Au final, si des pistes sont données pour la mise au point de nouveaux inhibiteurs pour Grb2 SH2, le protocole employé n'a pu être validé, si bien que contrairement à la dynamique moléculaire nous ne sommes pas en mesure de soumettre des propositions précises afin d'améliorer le protocole de screening virtuel qui a été mis en place. Toutefois, une analyse des différentes techniques de docking disponibles donne un aperçu des aspects théoriques et techniques dont l'amélioration pourrait être la plus profitable.

Ce travail peut bien sûr être soumis à plusieurs critiques. Si on peut estimer qu'il représente un certain avancement dans la compréhension du mécanisme d'inhibition du récepteur de Grb2 SH2, tout en contribuant à l'amélioration des approches utiles à une telle étude, il est aussi clair que de nombreuses possibilités restent ouvertes.

En ce qui concerne la dynamique moléculaire, si le rôle des molécules d'eau constituant l'environnement biologique du système a été mis en évidence pour un ligand précis avec Grb2 SH2, peut-on procéder à une quelconque généralisation ? La réponse est négative, en l'absence de données ultérieures issues de simulations supplémentaires ou bien d'expérimentations. Nous nous contentons d'indiquer qu'en l'absence de connaissances précises préalables sur les effets de solvant possibles d'un système protéine / ligand, il convient par prudence de modéliser le solvant dans la limite de ce que les méthodes employées permettent.

Le protocole de docking flexible n'est également pas validé à l'heure actuelle. Dans ces conditions, nous ne pouvons pas qualifier la valeur des résultats obtenus. Mais là encore, un protocole a été clairement établi, et les résultats sont disponibles : il s'agit de la première étape d'un travail de longue haleine qui s'intègre dans le développement de la plateforme VSM-G.

Sur un plan personnel, cette étude s'est avérée au final bien plus complexe qu'elle ne le paraissait au premier abord. Si j'étais conscient au début de mon doctorat que les techniques de la chimie informatique ne sauraient se limiter à lancer des calculs et à analyser leurs résultats, et implique une remise en question permanente des modèles, des techniques et des méthodes d'analyse, j'ai fait preuve d'un certain optimisme. Je souhaite que le lecteur de ce document, s'il n'en était pas déjà convaincu, aura pu percevoir la complexité mais aussi la richesse et le potentiel actuel des méthodes dérivées de la modélisation moléculaire.

Publications

Article 1

Leroux V, Gresh N, Liu WQ, Garbay C, Maigret B
Role of water molecules for binding inhibitors in the SH2 domain of Grb2 : a molecular dynamics study.
J.Mol.Struct. **806** (2007) 51-66
DOI : 10.1016/j.theochem.2006.11.010

Article 2

Beautrait A, Leroux V, Chavent M, Ghemtio L, Desvignes MD, Smaïl-Tabbone M, Cai W, Shao X, Moreau G, Bladon P, Yao J, Maigret B
Multiple-step virtual screening using VSM-G : overview and validation of fast geometrical matching enrichment
J.Mol.Model. **14** (2008) 135-148
DOI : 10.1007/s00894-007-0257-9

Alexandre Beautrait, en tant que programmeur de VSM-G, est naturellement le premier auteur de cet article. En dehors de la rédaction avec Alexandre, j'ai pour ma part effectué les calculs utilisant GOLD, ai fourni des codes de bas niveau, en particulier pour la conversion de formats, ainsi que le programme de création de "barres d'énergie" déjà utilisé dans l'article 1. J'ai également conçu et mis en place l'ensemble des techniques d'analyse / validation des résultats.

L'ancienne révision de ce document, disponible au moment de la soutenance de thèse, présentait une version antérieure de cet article, alors en cours de soumission. L'article avait pour titre :

VSM-G : the Virtual Screening Manager for computational Grids. Example of use for the identification of putative liver X receptor ligands.

Cette version est identique à l'article finalement publié en ce qui concerne la description de l'outil VSM-G, mais différait au niveau du protocole de screening utilisé pour valider celui-ci. Cette validation, moins étendue, reposait sur la "récupération" par VSM-G de 3 ligands connus de LXR au sein d'une base de "leurres". Elle a été abandonnée car elle ne permettait pas de mettre en évidence l'efficacité de MSSH/SHEF au sein de VSM-G dans un contexte de criblage haut-débit "en aveugle". En particulier, la similarité des leurres avec les références a été jugée trop faible pour convaincre des performances de filtrage de VSM-G. Cette ancienne version de l'article est disponible sur demande.

Article 3

Leroux V, Maigret B
Should structure-based virtual screening techniques be used more extensively in modern drug discovery ?
Computers and Applied Chemistry **24** (2007) 1-10

Role of water molecules for binding inhibitors in the SH2 domain of Grb2: a molecular dynamics study

Vincent Leroux [1], Nohad Gresh [2], Wang-Qing Liu [2], Christiane Garbay [2], Bernard Maigret [1,*]

1: Université Henri Poincaré – Nancy I, UMR CNRS / UHP 7565, groupe eDAM, BP 239, 54506 Vandœuvre-les-Nancy Cedex, France

2: Université René Descartes – Paris V, UFR Biomédicale, Laboratoire de Pharmacochimie Moléculaire et Cellulaire, 75006 Paris, France; Inserm, U648, 75006 Paris, France; CNRS, FRE 2718, 75006 Paris, France

* Corresponding author: Bernard.Maigret@edam.uhp-nancy.fr, http://www.edam.uhp-nancy.fr

Keywords
Grb2 SH2, Molecular Dynamics, solvent effects, water bridging, ligand binding, drug design.

Abstract
Growth factor receptor-bound protein 2 (Grb2) plays an essential role in the Ras-MAPK signalling pathway which is an important target for anti-cancer drug design. The precise mechanisms by which effective and selective ligands can bind to the Src homology 2 (SH2) domain of Grb2 and interrupt the signalling pathway are not fully understood. We report in this paper the results of molecular dynamics simulations of the Grb2 SH2 domain structure derived from X-ray in water solution and without constraints. The protein was complexed with two reference molecules: one is a potent and extensively studied inhibitor, while the other was shown to have no affinity for the Grb2 SH2 domain, in contradiction with previously performed *in vacuo* simulations. Analysis of the MD trajectories for the two complexes reveals interesting features which may explain the stability of the complex obtained for the first molecule versus the poor binding of the other ligand. It is shown that water plays a critical role in the stability of these complexes, and it is proposed to consider this solvent behavior in forthcoming structure-based drug design strategies. In this respect structural details are given concerning water molecules reported to be stabilized between the active ligand and the protein receptor. We have also identified which residues from Grb2 SH2 and the two ligands could be involved in destabilizing interactions with bulk solvent. Finally, we propose guidelines for optimizing both ligands by preventing such unwanted effects.

Abbreviations
EGF-R : epidermal growth factor receptor
Grb2 : growth factor receptor bound protein 2
IC_{50} : concentration required for 50% inhibition
MAPK : mitogen-activated protein kinase
MD : molecular dynamics
pTyr / pY : phosphotyrosine / phosphorylated tyrosine
Ras : oncogene isolated from rats and associated with sarcoma virus
RMSD : root mean square deviation
RTK : receptor with endowed tyrosine kinase activity
SH2 : Src homology 2
SH3 : Src homology 3
Shc : Src homology 2 domain containing
Sos : son of sevenless protein
Src : sarcoma viral oncogene homolog

Introduction

Several signalling and adaptor proteins have an SH2 (Src homology 2) domain which regulates the activity and specificity of kinases and binds phosphotyrosine (pTyr) residues of growth factor receptors. [1, 2] Such a binding triggers a cascade of metabolic events resulting into nuclear transcription. [3] SH2 domains could thus constitute an essential target for the design of small molecules that, upon competing with the binding of pTyr residues, would act as regulators of signal transduction. An important pharmaceutical interest arises from the fact that over-expression of such proteins, corresponding to a deregulation of the pathway and thus of cell-cycle progression, is taking part in processes of cellular hyper-proliferation and cancer development. [4-7]

One of the most important SH2 domain-encompassing proteins is Grb2, which is an adaptor composed of one SH2 domain linking two SH3 domains. [8-10] It has a significant activity as a linker with numerous receptors of the cell membrane [11] also involved in fibroblast [12] and neoplastic transformations [13]. Complexed with Sos, the exchange factor of Ras, the SH2 domain of Grb2 binds directly to auto-phosphorylated RTKs such as EGF-R [14] or via Shc [15], then activating the Ras-MAPK signalling pathway. [16, 17] This pathway is essential for cell growth and differentiation, and its mutations are amongst the most frequently observed ones in human cancers. [18] Anarchic cell proliferation characterizing acute myeloid leukaemia [19], breast, ovarian and prostate cancers [20], was shown to be related to over-expression of HER2/ErbB2, a truncated analog of the EGF receptor and Grb2 in the Ras pathway. Thus the search for Grb2 SH2 inhibitors, as for other compounds blocking the HER2 signalling, is a particularly interesting approach for anti-cancer drug design. [5-7, 21, 22]

The binding sites of SH2 domains are known to recognize very short sequences of amino acids, the most active part being composed of pTyr and three other residues C-terminal to it. Mutagenesis studies of small sequences of amino acids demonstrated each family of SH2 domains to recognize a well-defined sequence, namely pYVNV (the two valine residues not being crucial) in the case of Grb2. [23, 24] The binding mode of such a pattern is depicted by X-ray crystallography measurements [25, 26], and thus constituted the basis of several drug design studies for more efficient and specific pseudo-peptidic analogues [27, 28]. One compound, mAz-pY-(α-Me)pY-N-NH_2, hereby noted **ligand-1** (see **figure 1.a**), is a result of this research, and one of the most potent Grb2 SH2 inhibitors known to date. [29]

Ligand-1 targets three distinct binding sites on Grb2 SH2, respectively labeled **cavity-1**, **cavity-2** and **cavity-3** in the present work. Cavity-1 is the main pTyr binding pocket common to all SH2 domains. In Grb2 SH2, it is composed of residues R_{67} (Arg αA), R_{86} (Arg βB), S_{88} (Ser βB), S_{90} (Ser BC) and S_{96} (Ser βC). This Grb2 SH2 residue notation corresponds to the residue numbering in Grb2 as reported in the 1JYQ PDB structure, alternate between brackets is as proposed by Eck [30] for defining the α helix-ß sheet succession ßαßßßßßαß seen in most SH2 domains. [31] Cavity-2 is composed of K_{109} (Lys βD), L_{120} (Leu βE) and W_{121} (Trp EF). While such an hydrophobic binding site is present in all SH2 domains, in the case of Grb2, the specific W_{121} (Trp EF) residue restrains its size and forces ligands to adopt a β-turn conformation [32]. Peptidic and pseudo-peptidic ligands can adopt such a required conformation with the presence of an Asn_{+2} residue [33] (the +2 in subscript indicates that the Asn is 2 residues C-terminal to the main $pTyr_0$). Cavities 1 and 2 are both targeted by the natural Grb2 SH2 ligand – peptide sequence pYVNV – and it should be the case as well for potential ligands. [34] Cavity-3 constitutes a second pTyr binding pocket specific to Grb2 and, to our knowledge, only targeted by ligand-1. It is composed of residues S_{141} (Ser BG), R_{142} (Arg BG) and N_{143} (Asn BG).

The X-Ray structure of the SH2 domain of Grb2 in complex with ligand-1 is available in the Protein Data Bank [35] (PDB code is 1JYQ). Typical peptidic or pseudo-peptidic ligands for Grb2 SH2 [26, 34, 36], including the natural recognized sequences pYXNX, have a similar binding mode to Grb2 SH2 than ligand-1: indeed, all of those ligands present a $pTyr_0$ residue bound to cavity-1 and an Asn_{+2} residue bound to cavity-2. The binding of Grb2 SH2 ligands is mainly based on these two interactions. Ligand-1 specificity relies on the optimization of the mAz_{-1} and (α-Me)$pTyr_{+1}$ residues, which does not alter the binding of $pTyr_0$ and Asn_{+2}. Consequently, it can be said that proposals of novel drug design strategies for Grb2 SH2 inhibition deriving from the study of ligand-1 would also be of interest when applied to most Grb2 SH2 ligands.

Additional work was conducted in order to further increase the affinity provided by ligand-1. Molecular modelling studies [37] were performed in order to design novel compounds by optimizing the corresponding interactions, starting from the published X-ray structure of ligand-1 complex [26] (PDB entry 1JYQ). A new ligand, hereby noted **ligand-2** (**figure 1.b**), was thus designed and synthesized, and its efficiency was measured using ELISA [38] tests. However, contrary to *in vacuo* MD predictions, ligand-2 was shown to be experimentally unable to inhibit Grb2. As these calculations were made without taking any water molecule into account, assuming that solvent interactions with the protein-ligand system were not crucial, we made the hypothesis that the discrepancy between simulation and experiment could originate from that point. Recent papers about a different SH2 domain raised our interest towards that direction: in the case of Src SH2 in its complex with the peptide pYIpYV [39], the $pTyr_{+2}$ residue appears bound more to the water network around its position than to the protein, indicating that water could play a significant role in the binding process. Furthermore, an analysis of available thermodynamical data for the SH2 domain of Src complexed by the pYEEI sequence suggested that solvent effects could contribute by as much as 25% to the total binding free energy. [40]

In the case of the complexes of the Grb2 protein with ligands, high-resolution X-ray crystallographic data [26, 34, 36] also reveal that a limited number of water molecules are present at the protein-ligand interface, some of them buried inside cavity-1 with the $pTyr_0$ residue. The purpose of the present study is to contribute to unravel the possible role of the solvent and of these water molecules in particular, in the binding of inhibitors to Grb2 SH2 and to provide insight towards improved inhibitor design. Therefore, in order to account for solvent effects on binding, we have performed MD studies of the complexes of both ligand-1 and ligand-2 docked to Grb2 SH2 with the interface waters explicitly included, the complex being immersed in a water box with periodic boundary conditions. Applying the same procedure to ligands 1 and 2 should indicate if water could affect differently the latter's binding modes or energies.

Figure 1
Chemical structures of ligand-1 and ligand-2

(a) ligand-1

(b) ligand-2

Material and Methods

Design and Synthesis

The design and synthesis of the pseudo-peptidic linear ligand-1 were already described in previous works. [6, 29] In its complex with Grb2 SH2, as observed in the crystal structures, ligand-1 adopts a β I-turn conformation. β II-turn mimetics have been largely described in the literature, contrary to β-I turn mimetics. [41] In such a turn structure, the oxygen of the -1 carbonyl group forms an H-bond with the backbone NH of the +2 residue. Eguchi *et al.* have developed a bicyclic structure, tetrahydro-2H-pyrazino[1,2-α]pyrimidines-4,7-diones, as a β I-turn mimetic. The advantage in this structure is that all the side chains of a four residues peptide are reserved, which is uncommon with such mimetics. [42, 43] For this reason this bicycle structure was adopted as the starting point for the development of non-peptidic Grb2 SH2 inhibitors, first leading to a mimetic of Ac-pTyr-(αMe)pTyr-Asn-NH-$(CH_2)_3$-(1-naphtyl). In the proposed molecule (ligand-2), both phosphate groups of pTyr and (αMe)pTyr are replaced by phosphonate groups, which are stable to phosphatases.

The pTyr mimetic building block was prepared from 2-(4-bromo)phenyl acetic acid (**figure 2**). The preparation of the (αMe)pTyr mimetic, a β^2 amino acid was made by the method described by Liu *et al*. [44] The C-terminal hydrophobic building block, 2-(1-naphtyl)propylamine, was prepared by the method described by Furet *et al*. [45] The bicyclic molecule was synthesized following the method described by Eguchi *et al.* starting from a NovaSyn TG hydroxyl resin. The diethyl 2-bromoacetal was condensed with the hydroxyl resin and then alkylated by 3-(1-naphthyl)propylamine. Fmoc-Asn(Trt)-OH, suitably protected Fmoc-β^2-aminoacid and pTyr mimetic building block were introduced successively by HATU / HOAt / DIEA coupling and 20% piperidine Fmod-deprotection. The peptidyl resin was then freed from the resin by formic acid catalyzed hydrolysis of the acetal group to form the aldehyde function, which cyclized spontaneously to give the bicyclic molecule. The final product was obtained by treating the resin-free molecule with TMSI to hydrolyze the phosphonate protections (**figure 3**).

Computational simulations

We have used the 1JYQ PDB structure [26] as a starting point. It corresponds to the SH2 domain of Grb2 complexed with ligand-1. Since it is a dimer, we only retained one monomer (chain identifier B; chain A and its ligand was removed). Hydrogen atoms were then added according to the expected charge distribution of amino acids at pH = 7. The particular protonation state of H_{107} (His βD) (present in the binding site) was taken according to the one found in the NMR structure of the Grb2 protein complexed with a Shc-derived peptide (PDB code 1QG1). [46] This histidine protonation state appears optimal for intramolecular interactions with neighbor residues. All the water molecules from X-ray data that were distant by more than 5 Å from the protein or the ligand were suppressed. We thus obtained a set of 16 water molecules located in or near the protein-ligand binding site and these were labeled specifically in order to follow their individual trajectories during the MD simulations. The complex (1551 atoms from 96 residues + 87 ligand-1 atoms) was next immersed at the center of a large cubic water box of 80 Å length (which contains approximately 16500 water molecules) to simulate the biological environment in a realistic way. TIP3 was the model chosen for all waters.

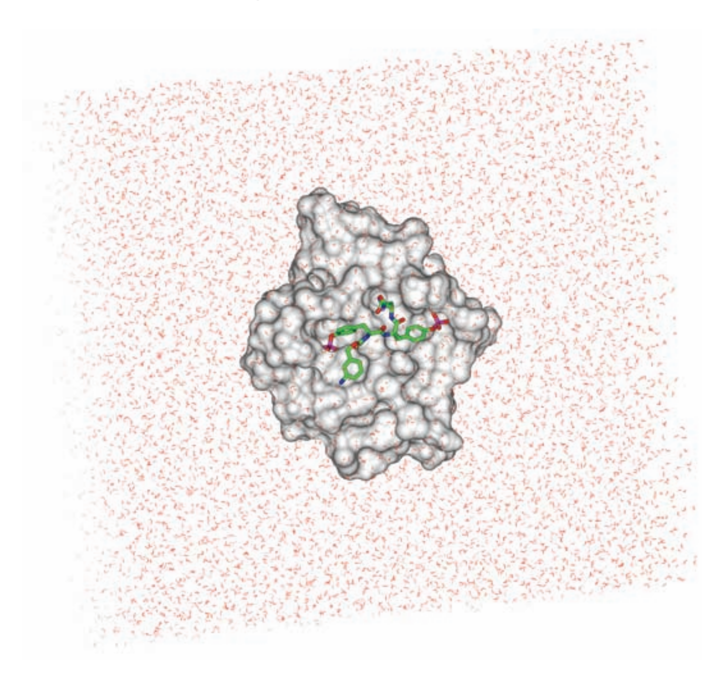
Figure 2
Preparation of the pTyr mimetic building block

Figure 3
Synthesis of ligand-2

The starting structure of the complex with ligand-2 was obtained by superimposing ligand-2 over ligand-1 using the two phosphorous atoms as superimposition points and then suppressing ligand-1. Two other systems were also preparated by simply removing all water molecules from the previously described models, in order to perform again the *in vacuo* MD simulations for ligand-1 and ligand-2 complexes using the same parameters and force-field than the dynamics with explicit water. Another model was prepared from the ligand-1 complex *in vacuo* by removing ligand-1 and recreating the 80 Å waterbox. The initial conformation of Grb2 SH2 is not initially relevant in such a system because it includes the induced fit effects upon ligand binding while the ligand has been removed. Our first goal here is to retrieve the unbound Grb2 SH2 conformation that is known experimentally from the complexed structures, which would be an indication of the correctness of these models. Besides, after this conformational change, this MD simulation will provide interesting information regarding the effects of solvent on receptor residues.

To generate trajectories of the complexes we used the NAMD program [47] coupled with the adapted CHARMM22 force-field. Before running the calculations, appropriate counter-ions were added replacing water molecules at the boundaries of the water box. **Figure 4** represents the starting structure as modelled for the ligand-1 complex MD simulation. Periodic boundary conditions were used for the three models including the 80 Å waterbox, and all models were processed at constant pressure of 1 atm and at a constant temperature of 300 K using Langevin piston. No constraints were applied, except for the O-H water bonds that are kept rigid in the simulations with water. Switching cut-off was set from 8 to 10 Å.

Figure 4
Modelled starting structure for the solvated ligand-1 complex MD simulation.



The calculations were performed on an SGI 3800 supercomputer and a Linux cluster. A minimization of 10000 steps using the conjugate gradients method was first performed. At the end of the minimization phase it was verified that the binding mode of ligand-1 and ligand-2 did not change significantly by checking the hydrogen bonds. It was also verified that the binding mode of ligand-1 and ligand-2 was similar as predicted from previous works on ligand-2: we superimposed ligand-1 and ligand-2 complexes after minimization and noted that the positions of the three functional groups of both ligands (the two phosphate / phosphone groups and the Asn_{+2} residue) relative to the Grb2 SH2 receptor matched. 10 ps of equilibration were then performed, giving the initial coordinates and velocities for the dynamics. The consistency of the structures was verified another time at this point. The time steps of the dynamics were set at 1 fs. The total simulated time for each of the five trajectories (ligand-1 and ligand-2 complexes in 80 Å waterbox, ligand-1 and ligand-2 complexes *in vacuo*, uncomplexed Grb2 SH2 in 80 Å waterbox) is 2 ns. A conformation was recorded each 5 ps, except for the uncomplexed Grb2 SH2 MD (10 ps), giving 400 conformations per trajectory (200 for uncomplexed Grb2 SH2). During the dynamics, it was verified that Grb2 SH2 did not undergo abnormal structural changes, that the size of the unit cell did not fluctuate much, and that the counter-ions always stayed out of reach (> 10 Å) from Grb2 SH2 and thus did not directly influence its dynamics.

Analysis of the MD trajectories

Analysis of the MD data was performed either using NAMD or InsightII (Accelrys). In the latter case we had to make an in-house conversion program to obtain .arc format trajectory files from the .dcd format outputted by NAMD. We also had to truncate the resulting .arc files that were too big (> 2 Gb for the solvated complex trajectories) for InsightII to load them. Consequently, the snapshots of ligands conformations, obtained with the protein represented with its Connolly surface (and colored upon its distance to the ligand using an InsightII script) were generated from .arc files with all water molecules suppressed. For the generation of cluster graphs only the ligands were retained. For the analysis of the position of water molecules regarding the protein-ligand interaction, we programmed a filter that works as follows. (i) In the case of the ligand-1 complex, all waters retained from the X-ray structure are conserved. (ii) Only the water molecules from the

waterbox that are distant by less than 5 Å from both the protein and the ligand during a continuous period of at least 50 ps of simulation are maintained. (iii) Only the protein residues distant by less than 5 Å to the ligand at any time are maintained. (iv) Of course, all ligand atoms are maintained. Please note that such truncated trajectory files are only used for ”graphical” analysis using InsightII; all other analyses (with NAMD) make use of the original .dcd files.

We used cluster graphs to study the stability of both ligand-1 and ligand-2 conformational behavior during the MD simulations. Such graphs represent the RMSD between the different conformations obtained during the simulation. The x and y axis correspond to the simulation time, and each point represents the RMSD between conformation at timestep x and timestep y, thus on the diagonal we have RMSD = 0 Å, and every cluster graph is symmetric regarding its diagonal. With the color scale used, black, white, blue, pink and red correspond to slight variations that indicate normal fluctuations at 300K, while green, cyan, yellow (then black again) highlight more or less important conformational changes. The cluster graphs are obtained by reprocessing InsightII-generated graphics. They only represent the absolute stability of the ligands conformation over time and are not directly related to its stability regarding the receptor nor to the binding energies. They are however very useful for identifying the most relevant ligand conformational changes that occur during the dynamics.

NAMD has a feature described in the manual as ”pair interactions” that changes the energy output to just a part of it corresponding to the interactions between two user-defined atom groups. It is possible to extract such values multiple times (with different group definitions) by reprocessing the trajectory file in a series of ”forced” runs at 0 K (this procedure is described in detail in the NAMD manual). Using this method we computed the interactions between various groups: Grb2 SH2, the ligands, residues of the Grb2 SH2 receptor, specific parts of the ligands, water molecules. The calculated energy values solely represent the expression of the interaction potential between specific non-bonded groups of atoms; this is just the sum of the van der Waals and Coulomb interactions. Entropy is not evaluated at all. Such values which we designate as interaction energy or binding energy are not the most representative of the interaction between the protein and the ligand; the free energy of binding is. However, our goal is not to compute binding energy values precisely, but rather to gather data that is able to clearly point out the differences between ligand-1 and ligand-2 binding.

In order to obtain an intuitive representation of that data, programs were made in order to build automatically the files describing the atom groups as well as the corresponding NAMD configuration files, and to start the NAMD jobs. Another program was made for converting a series of time / energy values such as those obtained with the aforementioned calculations into graphic files containing energy bars (with time as x-axis variable) using a black and white or a color scale from $E \geq 0$ (white) to $E \leq$ [user-defined negative value] (black). This procedure for representing the distribution of the binding interactions will be used systematically in our group to obtain a quick representation of the energetical repartition of the interactions between bio-molecules, and to monitor their fluctuation during MD simulations. We also plan to implement it in VMD and its InterSurf plugin [48]; so that protein binding sites can be spotted easily, and the corresponding interactions clearly represented graphically during dynamics analysis.

Results

Experimental measurements

The measurement of the affinity of ligand-1 for Grb2 SH2 has already been presented. [29] According to the ELISA experiments, while ligand-1 exhibits a very strong inhibitory activity on Grb2 SH2 signalling, with an IC_{50} of 11 ± 1 nM, ligand-2 presents no such effect ($IC_{50} > 10^{-4}$ M).

Evolution of the conformation of ligand-1 and ligand-2 complexes

The behaviors of ligand-1 and ligand-2 are entirely different during the MD simulations. The conformational variations of the ligands are depicted as cluster graphs in **figure 5**. Structures of both complexes at different times are shown in **figure 6**.

Only limited variations around the initial X-ray conformation take place with the ligand-1 complex trajectory, conserving the main interactions of the two phosphate groups. The most persistent hydrogen bonds are those occurring between the R_{86} (Arg βB) guanidinium group and the $pTyr_0$ phosphate. This phosphate is held in cavity-1 by a network of H-bonds donated by R_{86} (Arg βB), S_{88} (Ser βB), E_{89} (Glu BC) and S_{96} (Ser βC). The $pTyr_{+1}$ phosphate is anchored less deeply in cavity-3 through H-bonds mostly involving R_{142} (Arg BG) guanidinium and S_{141} (Ser BG) hydroxyl groups. The linker moiety between the two phosphate groups is maintained on the binding surface by H-bonds involving mainly the guanidinium group of R_{67} (Arg αA). The stacking of the mAz_{-1} aromatic ring with R_{67} (Arg αA) that was expected from *in vacuo* simulations appears unstable, flipping between its initial parallel conformation and an anti-parallel conformation oriented towards the solvent.

By contrast, the ligand-2 complex trajectory presents several important conformational changes as illustrated by cluster graph analysis (**figure 5.b**). During the simulation, the two phosphone groups present a clear tendency to escape from cavities 1 and 3, increasing the number of hydrogen bonds with the solvent at the expense of those with the protein. Moreover, the $naphthalene_{-1}$ moiety has immediately positioned itself in the bulk instead of retaining a stacking interaction with R_{67} (Arg αA) as expected from *in vacuo* simulations. Compared to the complex of ligand-1, conformational changes of both ligand-2 and Grb2 SH2 occurred so that several side chains of the protein binding site, initially interacting with ligand-2, have formed intra-molecular H-bonds, or H-bonds with waters, instead of the original ones. In cavity-3, this is especially notable for the S_{141} (Ser BG) residue which only interacts significantly with the $phosphone_{+1}$ group during the first 600 ps. Conversely, the $phosphone_{+1}$ and the R_{142} (Arg BG) guanidinium moved altogether towards the bulk, conserving their salt-bridge interaction. At t = 1 ns the $phosphone_{+1}$ appears only bound to this residue, as shown on **figure 6**. In cavity-1, S_{96} (Ser βC) hydroxyl group no longer interacts with the $phosphone_0$ after 200 ps, and the one of S_{90} (Ser BC) moves rapidly between 4-6 Å from the phosphorous atom. In the same cavity, R_{67} (Arg αA) guanidinium fluctuates several times between 4-8 Å from the closest $phosphone_0$ oxygen. By contrast, R_{86} (Arg βB) and S_{88} (Ser βB) keeps interacting with the $phosphone_0$. The weak interaction with R_{67} (Arg αA) compared to ligand-1 binding is compensated by the formation of a stable salt bridge involving K_{109} (Lys βD) outside cavity-1, which is lost abruptly after 1800 ps in favor of bulk hydration, as shown at t = 2 ns on **figure 6**.

Figure 5

Cluster graphs of the (a) ligand-1 (top) and (b) ligand-2 (bottom) trajectories

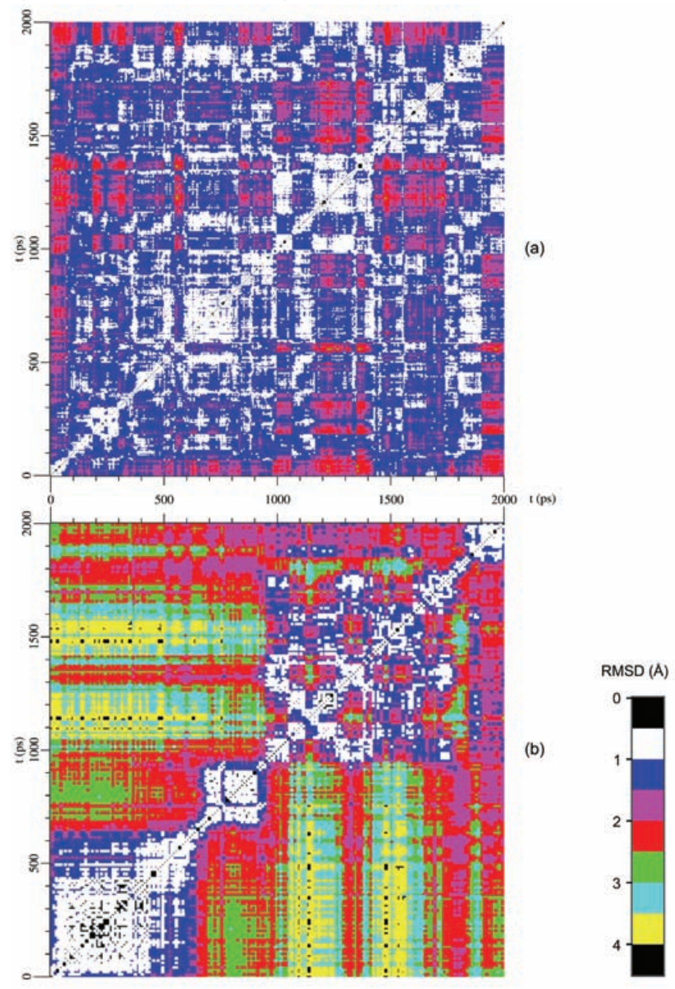
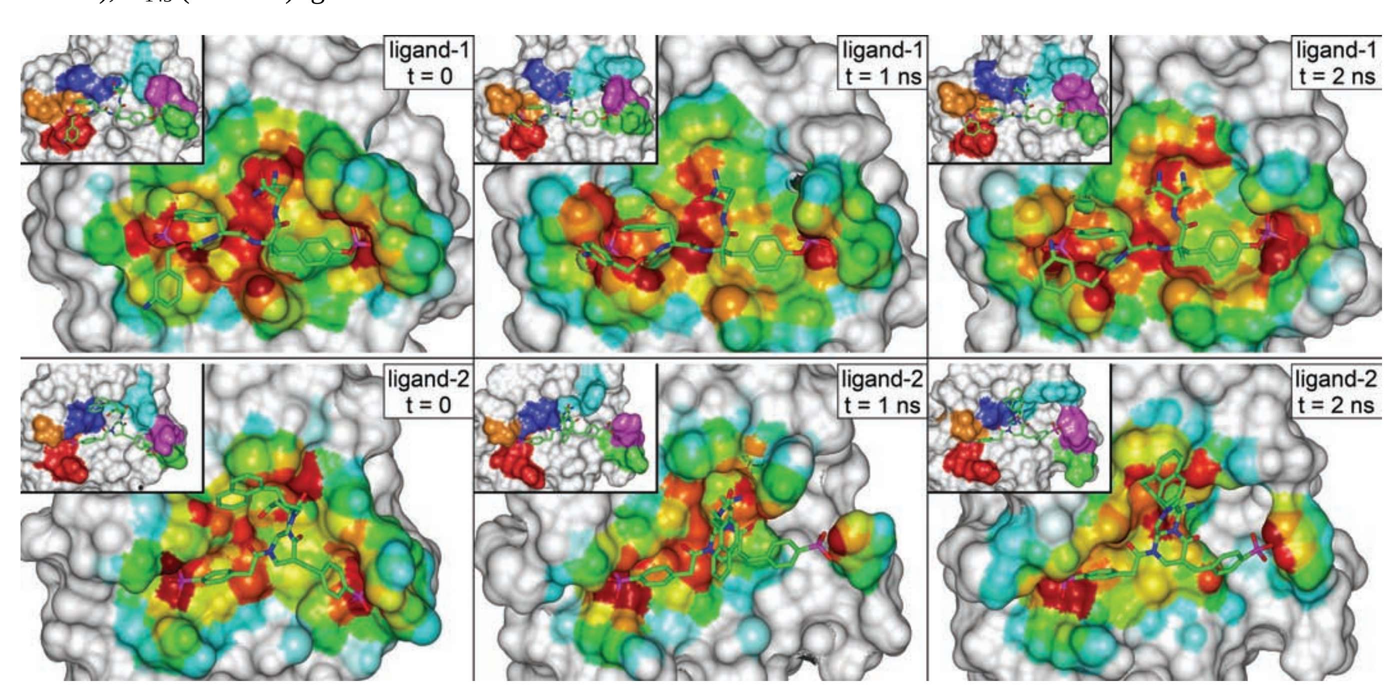


Figure 6

Snapshots of complexes of ligand-1 and ligand-2 during the MD simulations. The colors on the Grb2 surface on the main snapshots are related to the distance of the ligand to the surface. The colors on the smaller snapshots on the top red corners of the pictures are related to the main Grb2 SH2 interacting residues. Cavity-1: R_{67} (Arg αA), R_{86} (Arg βB) – red; S_{88} (Ser βB), S_{90} (Ser BC), S_{96} (Ser βC) – orange. Cavity-2: K_{109} (Lys βD) – dark blue; L_{120} (Leu βE), W_{121} (Trp EF) – light blue. Cavity-3: R_{142} (Arg BG) – purple; S_{141} (Ser BG), N_{143} (Asn BG): green.

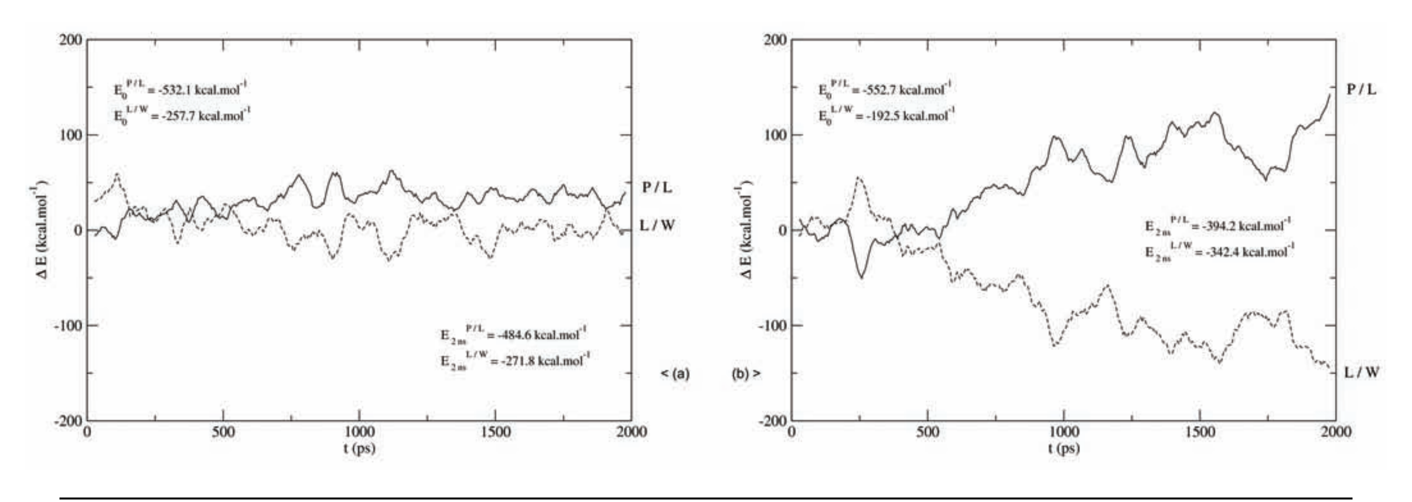


Evolution of the ligands binding energies

The decomposition of the interaction energy between ligand-1 residues and Grb2 SH2 indicates that $pTyr_0$ located in cavity-1 contributes to approximately 55% of the total protein-ligand binding energy, $pTyr_{+1}$ targeting cavity-3 to 40%, and the Asn_{+2} residue – which is required in order to have the specific Grb2 SH2 β-turn conformation – to 5%.

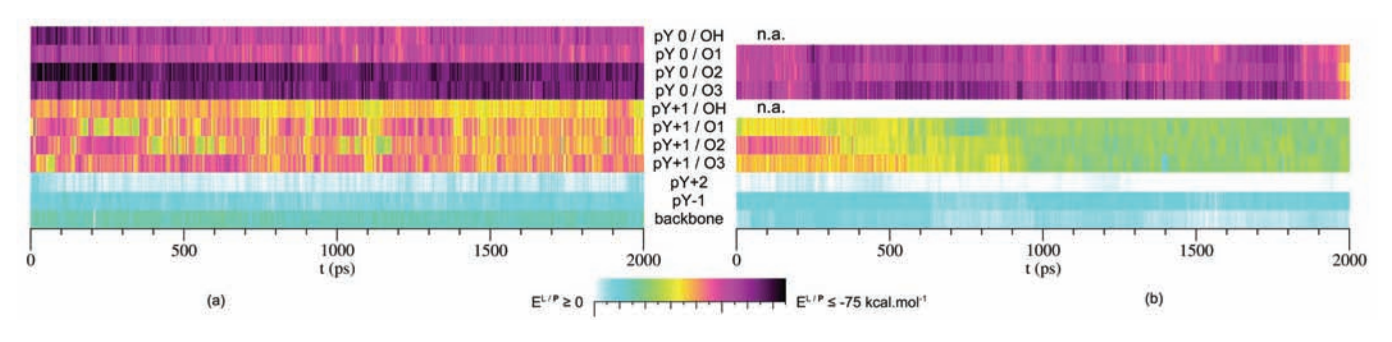
We have plotted the evolution of the binding energies between ligand-1 and ligand-2 on one hand, and the Grb2 SH2 domain and the water molecules on the other hand, in **figures 7.a and 7.b** respectively. These graphs clearly indicate that ligand-1 binding energy only undergoes acceptable fluctuations, as a loss of less than 10% of the interaction energy is observed after 2 ns. In an opposed way, a significant part of the initial ligand-2 interaction is lost in favor of a better interaction with the solvent. Thus, as already observed with the analysis of the conformations, the behaviors and ligand-1 and ligand-2 regarding the interaction energy with Grb2 SH2 differ completely. As expected from *in vacuo* simulations, ligand-2 in its initial conformation interacts more strongly with Grb2 SH2 than ligand-1 (interaction energy of -552.7 $kcal.mol^{-1}$ instead of -532.1 $kcal.mol^{-1}$), and less with water (-192.5 $kcal.mol^{-1}$ instead of -257.7 $kcal.mol^{-1}$). However, during the course of the MD, it appears that the ligand-2 interaction decreases significantly, to reach a level of interaction of the same order than its interaction with the solvent. This implies that ligand-2 can not be considered bound to Grb2 SH2 anymore, contrary to ligand-1.

Figure 7
Variations of the interaction energy between the Grb2 SH2 protein and the ligand (P / L, plain curves), and between the ligand and water (L / W, dotted curves). (a) corresponds to the ligand-1 complex (left), (b) to the ligand-2 complex (right).



In order to account more precisely for the binding of ligand-1 and ligand-2 on Grb2 SH2, the interactions of specific ligand groups were next computed. We retained (i) the 4 phosphate oxygens of the two phosphates of ligand-1 / the 3 phosphone oxygens on the corresponding phosphones of ligand-2, which are the only atoms of these groups that make significant favorable interactions (ii) the Asn_{+2} side-chain bound to cavity-2, (iii) the aromatic group in -1 position that was modelled in order to make stacking interactions, and (iv) the backbones of the ligands, peptidic in the case of ligand-1, non-peptidic and cyclic in the case of ligand-2. These results are presented as energy bars in **figure 8.a** for ligand-1 and in **figure 8.b** for ligand-2.

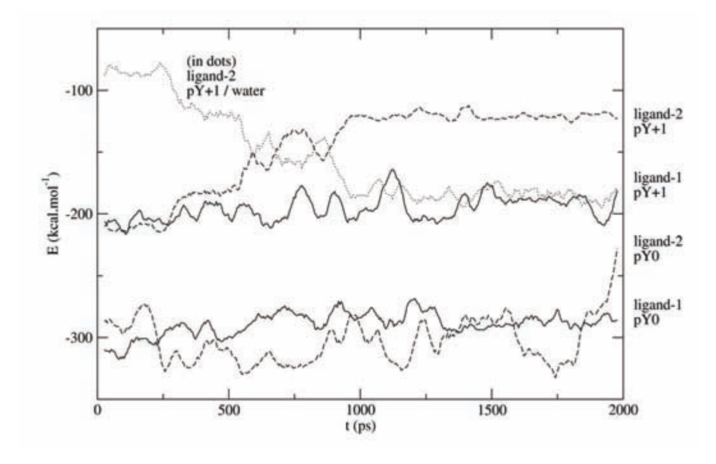
Figure 8
Interaction energy between various ligands atoms / groups and Grb2 SH2.
(a) ligand-1 complex (left); (b) ligand-2 complex (right).



The ligand-1 $pTyr_0$ phosphate oxygens, in agreement with the structural analysis, are bound more tightly to cavity-1 than their $pTyr_{+1}$ counterparts in cavity-3. In both cases, it appears that the non-terminal phosphate oxygens make interactions equivalent to those of the terminal ones, which is surprising because only the latter are H-bonded to Grb2 SH2, as predicted by the X-ray structure. In the case of ligand-2, as expected, the oxygens of both phosphones are equivalent, and their interaction with Grb2 SH2 is similar in magnitude to the one of ligand-1 phosphate oxygens. Regarding the residues in -1 position, both the ligand-1 mAz_{-1} and the ligand-2 $naphthalene_{-1}$ contributions to the binding energy are negligible during the whole 2 ns of trajectories, which confirms that these groups do not make stacking interactions with Grb2 SH2, whereas they clearly interact with water. Weak favorable interactions observed with mAz_{-1} are solely due to the presence of the terminal NH_2 group, which however is not H-bonded to Grb2 SH2 as observed in the ligand-1 complex X-ray structure.

The binding energy curves of ligand-1 OPO_3^{2-} and ligand-2 PO_3^{2-} groups, including the unfavorable contributions of the phosphorous atoms, are presented on **figure 9**. **Figure 8.b** indicates that the interaction between the second ligand-2 phosphone and Grb2 SH2 decreases during the simulation. **Figure 9** shows more clearly that the discrepancy of ligand-2 binding observed in **figure 7.b** corresponds solely to this evolution of the +1 group binding. This interaction is initially approximately 100 kcal.mol^{-1} greater than the $phosphone_{+1}$ interaction with the solvent, a feature also observed in ligand-1. Between 600 ps and 900 ps, the former lost 50 kcal.mol^{-1} while the latter gained 50 kcal.mol^{-1}: ligand-2 is not bound to cavity-3 anymore. After 900 ps ligand-2 moves away from cavity-3 as it interacts significantly more with water than with Grb2. Interestingly, the sum of both interactions of the residue in +1 position is approximately constant during the whole simulation. Additionally, the interaction of the group in position 0 is abruptly lowered in magnitude at 1800 ps, as expected from the observation of the loss of the phosphone binding with the K_{109} (Lys βD) residue.

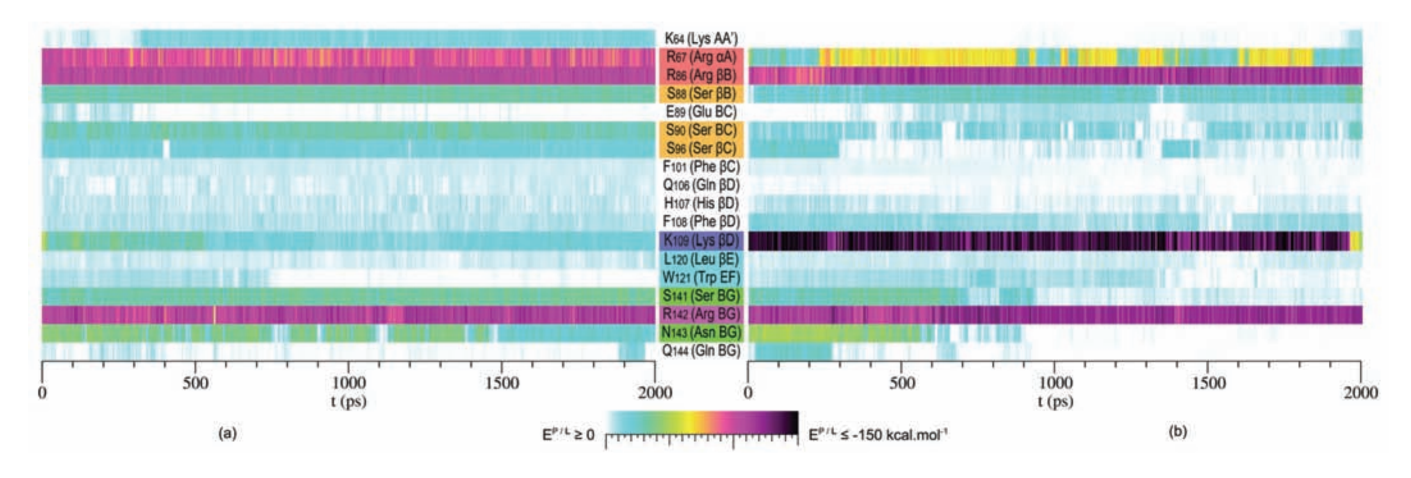
Figure 9
Interaction energy of the phosphate (ligand-1) / phosphone (ligand-2) group of the ligands.



Evolution of the protein residues binding energies

We also followed the variation of the interaction energy of all Grb2 SH2 residues with both ligands through each trajectory. This has permitted to establish the list of the Grb2 SH2 residues involved in ligand binding. As expected, residues R_{67} (Arg αA), R_{86} (Arg βB), S_{88} (Ser βB), S_{90} (Ser BC) and S_{96} (Ser βC) in cavity-1, K_{109} (Lys βD) in cavity-2 and S_{141} (Ser BG), R_{142} (Arg BG) and N_{143} (Asn BG), in cavity-3 have favorable ligand binding energies. The binding energy of residues L_{120} (Leu βE) and W_{121} (Trp EF) in cavity-2 is very low. In contradiction to what could be expected from the X-ray structure of the ligand-1 complex, the H_{107} (His βD) residue does not interact significantly with either ligand-1 or ligand-2. Additionally, K_{64} (Lys AA') makes minor interactions with both ligands. We will now focus on the data regarding the aforementioned interacting residues. The corresponding information is represented as energy bars in **figures 10.a and 10.b**.

Figure 10
Interaction energy between significant Grb2 SH2 residues and (a) ligand-1 (left) / (b) ligand-2 (right).
Background colors for residues names are the same as the ones employed in figure 6 small snapshots.



The variations of the binding energies of the involved Grb2 SH2 residues are small in the case of ligand-1, particularly with the residues forming cavity-1. The corresponding binding mode in cavity-1 involving R_{67} (Arg αA) and R_{86} (Arg βB) – but not K_{109} (Lys βD) – is less favorable energetically than the initial binding of ligand-2, which mostly involve R_{86} (Arg βB) and K_{109} (Lys βD). Indeed, the interactions with R_{67} (Arg αA) and the three cavity-1 serines are lower in the case of ligand-2 by 40 and 30 kcal.mol^{-1} respectively. This is compensated by the interaction with K_{109} (Lys βD), which fluctuates around -140 kcal.mol^{-1} as compared to -40 kcal.mol^{-1} with ligand-1. However, after 1800 ps of dynamics, the corresponding H-bonds are lost, which is correlated with the ligand-2 binding disruption observed in **figures 7.b and 9**. This leads to a conformation where R_{86} (Arg βB) is the only positively-charged residue making H-bonds with the phosphone$_0$, and where the interaction of K_{109} (Lys βD) with ligand-2 is now of the same order of magnitude as with ligand-1, translating a drop of approximately 100 kcal.mol^{-1}. Such a disruption of ligand-2 binding is not compensated by a strengthening of the remaining interactions. Thus the interaction of ligand-2 with cavity-1 is finally much weaker than the one of ligand-1.

Regarding the interactions of the residues forming cavity-3, the values obtained with ligand-2 give more details about the large energy loss discussed previously and observed in **figures 7.b, 8.b and 9**. The interactions with the S_{141} (Ser BG) and N_{143} (Asn BG) residues are lost, while the interaction with R_{142} (Arg BG) is conserved as observed with the analysis of the conformations. Thus, until the K_{109} (Lys βD) interaction in cavity-1 is lost at 1800 ps, the weakening of the binding energy between ligand-2 and Grb2 SH2 is linked to the loss of well-defined H-bonds, whereas these are maintained throughout the 2 ns of simulation in the case of ligand-1. Interestingly, during the simulation of the ligand-1 complex, the N_{143} (Asn BG) interaction with ligand-1 undergoes significant fluctuations without being lost. This could indicate that this interaction is prone to be disrupted upon taking solvent effects into account but is maintained with the help of the interactions with S_{141} (Ser BG) and R_{142} (Arg BG) which both remain stable. However, the ligand-2 case indicates that the main R_{142} (Arg BG) interaction could not be of enough amplitude so as to alone maintain the ligand in cavity-3, as it is also strongly interacting with water molecules. This leads to the ligand-2 $phosphone_{+1}$ moving towards the bulk simultaneously with R_{142} (Arg BG).

Behavior of specific water molecules along the trajectories

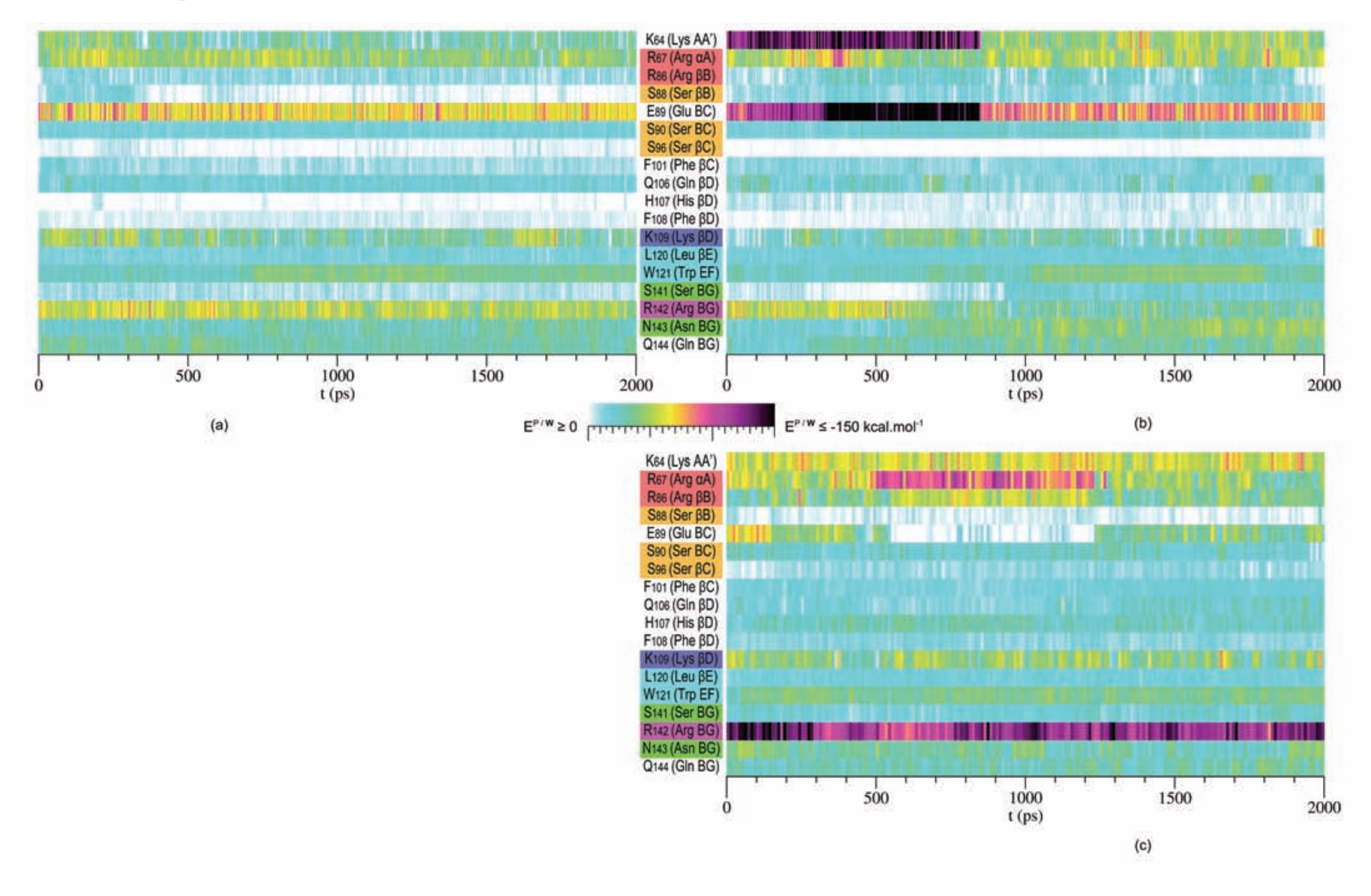
The positions of water molecules interposed between ligand-1 and Grb2 SH2 as depicted in the X-ray structure were tracked during the MD trajectory. It appeared that only one of those waters (HOH_{324} in 1JYQ, monomer B) did maintain itself interacting with both Grb2 SH2 and ligand-1 during the whole trajectory. HOH_{324} position in cavity-1 is fluctuating, and differs slightly from its conformation in the X-ray structure, where it is H-bonded to the terminal atoms of residues F_{95} (Phe βC), S_{96} (Ser βC) and K_{109} (Lys βD). During the MD simulation, HOH_{324} makes H-bonds mostly with the non-terminal $phosphate_0$ oxygen and the K_{109} (Lys βD) residue terminal nitrogen, but also with the terminal oxygens of the three cavity-1 serines and the A_{91} (Ala BC) residue. The other water molecules from X-ray quickly moved in the bulk, the most stable staying 100 ps and 350 ps. Throughout the trajectory, several water molecules from the water box retained stable interactions with Grb2 residues but did not interact significantly with ligand-1, with the exception of one water molecule that bridged S_{96} (Ser βC) and F_{108} (Phe βD), while remaining around 4 Å from the $phosphate_0$. Regarding water molecules exchanged at conserved positions, it is noted that in cavity-3, the $phosphate_{+1}$ is always H-bonded to at least two water molecules, thus defining a water H-bond network where the S_{141} (Ser BG) and N_{143} (Asn BG) residues are also involved.

Conversely, in the case of ligand-2, one water molecule remains in the vicinity of the $phosphone_0$ during almost 1 ns. This water makes H-bonds with the 3 $phosphone_0$ oxygens and with the terminal atoms of residues R_{86} (Arg βB), S_{88} (Ser βB) and S_{90} (Ser BC). It is thus located between the phosphone and the entrance of cavity-1. All other water molecules interacting with both the ligand and the protein exchange very rapidly. The averaged number of water molecules H-bonded to ligand-2 has increased from 4 to 9 as the simulation proceeded. This is correlated with the higher exposure of the ligand escaping from the active site.

Influence of bulk solvent on protein residues

As the tracking of a limited number of stable water molecules interacting both with the protein and the ligands could not alone account for the solvent effects on ligand binding, we measured the interaction energies between all protein residues and all water molecules. The results, restricted to the Grb2 SH2 receptor residues, are presented as energy bars in **figure 11.a** for the complex with ligand-1, in **figure 11.b** for the complex with ligand-2, and in **figure 11.c** for the uncomplexed Grb2 SH2.

Figure 11
Interaction energy between Grb2 SH2 residues and water.
(a) ligand-1 complex (top left)
(b) ligand-2 complex (top right)
(c) uncomplexed Grb2 SH2 (bottom)



With most of the residues on the ligand-1 complex the interactions with water slightly decrease during the simulation. In the case of the ligand-2 complex, at 1800 ps, as expected, the K_{109} (Lys βD) interaction with water increases as its interaction with ligand-2 is lowered. Additionally, stabilizations of S_{141} (Ser BG) and N_{143} (Asn BG) are observed. This corresponds to water molecules filling the gap in cavity-3 that results from ligand-2 escaping towards the bulk along with R_{142} (Arg BG). The interaction of the latter residue slightly decreases while its salt bridge with the phosphone$_{+1}$ is reinforced. This was not obvious as R_{142} (Arg BG) and phosphone$_{+1}$ move altogether away from cavity-3 and are therefore more exposed to the solvent. In uncomplexed Grb2 SH2, it is noted that R_{142} (Arg BG) is particularly hydrophilic as compared to R_{67} (Arg αA) and R_{86} (Arg βB) which are less exposed to the solvent in cavity-1. Interestingly, it is noted that the interactions of the E_{89} (Glu BC) residue with water differ with both complexes and with the uncomplexed Grb2 SH2. At the beginning of the simulation the E_{89} (Glu BC) makes much larger interactions with water in the case of the ligand-2 complex, with a difference of approximately 80 kcal.mol^{-1} with its interaction with water in the case of the ligand-1 complex. However, such an interaction with water goes to the same level as with the ligand-1 complex after 800 ps of simulation. The differences observed for the uncomplexed Grb2 SH2 correspond to the ligand not interacting with cavity-1 arginines, thus stabilizing their interaction with the glutamate. In all cases, at the end of the trajectories the Glu side-chain is displaced as to make a strong ionic interaction with the K_{64} (Lys AA') cationic side-chain; therefore only the Glu backbone is then located at the bottom of cavity-1.

Significant differences in the interactions of the charged Grb2 SH2 residues that are not directly involved in the protein-ligand binding are also observed. For example, K_{69} (Lys αA) is more solvent-exposed in ligand-1 complex than in ligand-2 complex, itself more exposed than in uncomplexed Grb2 SH2. The opposite is observed with K_{76} (Lys AB). The E_{72} (Glu αA) residue only interacts significantly with water molecules in the ligand-1 complex. The E_{87} (Glu βB) residue, similar to the R_{86} (Arg βB) residue involved in ligand binding, is much more solvent-exposed when Grb2 SH2 is ligand-bound. All those variations are correlated to induced fit effects upon ligand binding. Differences found in some cases between ligand-1 and ligand-2 could be the consequence of their adoption of different binding modes concerning their group in position 0 bound to cavity-1.

Force-field dependence

One aspect that could be of importance for explaining the initial discrepancy between experimental results and MD simulations is the force-field choice for the dynamics. All the simulations presented in this paper have been performed using CHARMM22 (and the NAMD program). The initial *in vacuo* simulations that prompted for ligand-2 synthesis and testing prior to this work were performed using CFF91 and the Discover program (Accelrys). In order to have a standardized protocol we recreated those simulations using CHARMM22. CFF91 and CHARMM22 *in vacuo* dynamics were in agreement, predicting that both ligand-1 and ligand-2 binding were stable on the Grb2 SH2 receptor. The MD simulations in water for ligand-1 and ligand-2 complexes were also performed using the CFF91 and AMBER force-fields, with Discover and parameters close to the ones employed with NAMD and CHARMM22. The results were again in agreement between the force-fields, predicting that when water is explicitly modelled ligand-1 is binds to Grb2 SH2 in a stable manner while ligand-2 does not. Structural analysis for both ligand-1 and ligand-2 solvated complexes showed conformations similar to those described previously for the CHARMM22 dynamics. These calculations allowed us to verify that for the system under investigation the choice of the force-field is not crucial (while the modelling of the solvent is).

Discussion

The importance of modelling water molecules

Until recently, upon modelling of protein-ligand complexes by docking strategies, solvent molecules were generally not taken into account, and were believed not to exert a significant effect on the binding. Inclusion of some water molecules in the molecular complex could have two opposite effects: (i) a favorable effect on the binding enthalpy, due to the formation of additional hydrogen bonds, and (ii) an unfavorable entropic effect due to the extraction of the water molecules from the bulk, restraining their available number of degrees of freedom. These two effects being in most cases considered of comparable impact, their sum is assumed to be negligible regarding the free energy of binding and have a similar magnitude in all structures considered. Additionally, a good inhibitor for a given receptor is usually supposed to remove water molecules initially present on the binding site. While generally this assumption is correct [49], it has been demonstrated that in some cases specific water molecules are to be considered as part of the receptor, and are better kept conserved upon ligand binding as they provide multiple stabilizing H-bonds that compensate for the absence of the entropic gain that could occur upon removal. [50]

The removal of water molecules from the modelled system was demonstrated here to be misleading in the case of Grb2 SH2, because only explicit inclusion of the water environment of two reference complexes was able to give results in agreement with experimental measurements. Moreover, the results of this study clearly indicate that the involvement of water molecules in the binding between a protein and a ligand can be separated in two classes: (i) favorable enthalpic interactions by a limited number of waters stabilized between the ligand and the protein surface, with some of them to be considered as part of the protein (namely those retained in both ligand-bound and free protein states, and reported as stable in MD simulations), and (ii) interactions with bulk solvent, which are unfavorable for ligand binding as they compete with it.

In the present study, it appeared that while the potency of ligand-1 could be enhanced by the contributions of a limited number of water molecules characterized by X-ray crystallography, the lack of affinity of ligand-2 was ascribed to destabilizing effects on the ligand caused directly by bulk solvent and/or to conformational changes of protein residues which occur only when solvent is explicitly modelled. Such effects could be correlated to specificities of the Grb2 SH2 system, as there are numerous examples of successful *in vacuo* drug design strategies indicating that most protein-ligand systems do not seem particularly sensitive to solvent effects such as those observed here.

Involvement of stabilized water molecules in the ligand-1 complex

Predicting how a given water molecule could affect the binding without resorting to complex calculations like the MD simulations performed in this study is not obvious because, as remarked by Dunitz [51], the entropic and enthalpic contributions to the free energy of binding are of almost the same magnitude for a hydrogen bond at 300 K. However, when a significant number of water molecules are found in several experimental structures interposed between the ligand and the receptor making multiple H-bond bridges, one should investigate further in order to determine if such waters would conserve their position in biological conditions, and therefore be structurally involved in ligand binding. On the opposite, if no such water is observed in X-ray structures, most probably no water molecule is involved in protein-ligand binding, because the inclusion of a water molecule from the solvent is more entropy-costly at 300 K than in X-ray conditions. With the X-ray structures of complexes of SH2 domains [26, 34, 36] we are in the former situation: one water molecule is observed in all cases at the bottom of cavity-1 forming H-bond bridges with one of the phosphate oxygens and the receptor. The X-ray structure of Grb2 SH2 complexed by ligand-1 [26] also embodies an additional number of bridging waters.

MD shows that the presence of such waters is not due to crystallization effects, since one of them appears to maintain its position during the simulation. In addition, upon exchange with waters from the bulk the positions of several other water molecules are conserved. This confirms that specific significant stabilizing water-mediated interactions occur at the interface between Grb2 SH2 and ligand-1. Thus the very strong affinity of ligand-1 compared to other known ligands could be assumed to rely not only on the targeting of cavity-3, but also on the involvement of those interfacial water molecules stabilizing the protein-ligand interaction by means of several additional H-bonds overcompensating for their loss of degrees of freedom upon binding. Indeed, this specificity, as well as cavity-3 targeting, could distinguish ligand-1 amongst other known inhibitors of Grb2 SH2 signalling, and thus those waters should be conserved upon modelling ligand-1 analogues. The present study only accounting for ligand-1 and ligand-2 binding, we cannot predict if they should also be conserved upon modelling other Grb2 SH2 ligands whose structure differ significantly. However, new methods should enable, starting from the structure of Grb2 SH2 including all possibly stabilized water molecules upon ligand binding, to sort those waters that should be conserved upon docking a given molecule. [52]

Disruption of ligand-2 binding by bulk water molecules

Independently from the stabilizing effect of a limited number of water molecules located at the protein-ligand interface, MD of both ligand-1 and ligand-2 complexes highlighted the destabilizing effects of bulk solvent. It first appears that bulk solvent can force specific ligand groups to keep enough solvent accessibility, resulting in a conformational change that could be unfavorable. In this study, this is observed with the ligand-1 mAz_{-1} group as well as with the ligand-2 $naphthalene_{-1}$. However, in the former case, this did not appear to lead to the disruption of the binding as all other groups did remain in their initial positions bound to cavities 1, 2 and 3. This observation is in agreement with biological assays, which demonstrated that the mutation of mAz_{-1} did not have a significant impact on ligand-1 affinity. [26]

More importantly, the analysis of MD data revealed that the lack of affinity of ligand-2 originates from the evolution of the binding of both phosphone groups, particularly the one targeting cavity-3. This does not seem to be caused by a weakness of the interaction between ligand-2 and cavity-3 residues, as it is initially of the same order of magnitude than the $phosphate_{+1}$ interaction of ligand-1, but rather in the hydrophilic nature of the very flexible arginine residue in cavity-3. This particular residue appears to have a clear tendency to move away from its serine and asparagine neighbors, thus changing the shape of cavity-3. Consequently, a ligand bound to cavity-3 has to maintain strong interactions with these residues, in order to force the arginine to keep its position in cavity-3 as observed in the ligand-1 complex X-ray structure. If this is not the case, there would be a strong probability that the arginine can move towards the solvent while maintaining a large interaction with a negatively-charged group of the ligand: this group would then follow the arginine and eventually be bound to water as strongly as to Grb2. This would correspond to a significant destabilizing conformational change of the ligand, coupled with a large drop of the binding energy. Such a transition occurs during the first nanosecond of the ligand-2 complex MD, and was not observed in the corresponding *in vacuo* MD simulation.

The fact that in ligand-2 the ligand-1 phosphates have been replaced by phosphones could explain why ligand-1 maintains its binding to Grb2 SH2 while ligand-2 does not. We demonstrated that in both cases all oxygen atoms bound to the phosphorus atom are making interactions of the same order with Grb2 SH2, and that as a whole the phosphate and phosphone groups are making interactions of similar strength when initially positioned with the same conformations regarding cavity-1 and cavity-3. In this regard, the only difference between those two groups relies in the geometrical distribution of the interactions: in three directions with the phosphone, in four directions with the phosphate. While this difference did not appear to lead to major consequences in *in vacuo* MD simulations, it seems to be crucial when solvent is modelled. With both cavity-1 and cavity-3, it appears that the four directions-conformation of a phosphate group leads to a more stable binding than the three directions-conformation obtained with ligand-2 phosphones. In this regard, the non-terminal oxygen of ligand-1 phosphate groups could constitute a conformational constraint which is very useful against bulk solvent destabilizing effects or potential unwanted conformational changes of specific protein residues caused by their interaction with the solvent.

Concerning cavity-1, with OPO_3^{2-} as a four-directional binding group, only deep binding in cavity-1 as performed by ligand-1 seems favorable. With the three-directional binding of PO_3^{2-}, a conformation locating ligand-2 partly outside cavity-1 is possible. The loss of binding observed at 1.8 ns at this level could have two causes: (i) the conformational change that is made on both the residue in +1 position targeting cavity-3 and the $naphthalene_{-1}$ residue, and (ii) the fact that the $phosphone_0$ group, not bound into cavity-1 like the ligand-1 $phosphate_0$, but rather at cavity-1 entrance, is much more exposed to the solvent, and thus to possible destabilizing effects. At this point it is only possible to affirm that at least one of these assumptions is correct; we cannot determine whether (i) is true, or (ii) is true, or both.

Proposals for modelling new Grb2 SH2 inhibitors

Useful additional information could be obtained by performing MD simulations on ligand-1 and ligand-2 analogues. Several strategies could be attempted in order to model such analogues. (i) The -1 group does not interact with Grb2 SH2 in the case of ligand-1 as well as with ligand-2: its suppression or replacement should be easily performed. (ii) The K_{109} (Lys βD) residue appears to be an interesting target for the binding of a negatively-charged residue on the ligand. This study indicates that one single doubly-charged residue such as a phosphate group can bind to either K_{109} (Lys βD) and R_{86} (Arg βB), or R_{67} (Arg αA) and R_{86} (Arg βB). Consequently, the design of analogues with two distinct charged groups, one targeting the two cavity-1 arginines and the other K_{109} (Lys βD), could lead to more potent compounds. (iii) It was shown that in cavity-3, while the interaction with the R_{142} (Arg BG) residue constitutes the main contribution of the interaction energy, the binding with the neighboring residues S_{141} (Ser BG) and N_{143} (Asn BG) have to be strong enough to constitute a constraint on R_{142} (Arg BG). Such a constraint is required as R_{142} (Arg BG) has a clear tendency to move towards the solvent, disrupting the conformation of ligands targeting cavity-3. This information should be taken into account when modelling such ligands.

Conclusions

As remarked by Ladbury [53, 54], leaving water out of a drug-design strategy reduces the likelihood that the strategy will be successful. This work shows that solvent effects can impact significantly the binding of ligands to the SH2 domain of Grb2, while the force-field choice does not. Consequently, not taking solvent into account with such a system could turn *de novo* design of active ligands to be unpredictive. [55] This is also in line with the conclusions of previous studies bearing on the SH2 domain of Src [39, 40], which indicated that solvent effects could intervene at the pTyr binding site, common to all SH2 domains, as well as in specificity zones. The present MD simulations confirm this behavior of water concerning the SH2 domain of Grb2, but also highlight two critical issues: (i) the involvement of a limited number of water molecules bridging the ligand and the receptor in positions persistently occupied during the simulations, and, more importantly, (ii) the interaction between bulk water and specific protein and ligand residues, which appeared to lead to disruption of ligand binding in one case. Therefore, this work suggests that the Grb2 SH2 binding site could be partly hydrated upon ligand binding, as was observed with other systems [56, 57], and gives precise information for the prospective inclusion of discrete water molecules in Grb2 SH2 models as input in library-screening methods that use structural data. In this regard, future studies will use the GOLD docking program [58, 59], for which the structural water issue was recently addressed [52] and, as a more accurate step in screening, the SIBFA molecular mechanics procedure [60, 61], where the additional effects of bulk solvation can be accounted for by a Continuum reaction field [62]. Moreover, the knowledge of the potential solvent effects on the binding of ligands on the SH2 domain of Grb2, as well as the MD protocol that can be used to unravel them, fulfills the requirements for designing new potent ”water-friendly” ligands. The synthesis and activity measurement of analogues of ligand-1, ligand-2, and a new class of Grb2 SH2 ligands [63] (if possible) will therefore be performed shortly.

Acknowledgements. The MD computations were performed on the computers of the Centre d’Informatique de l'Enseignement Supérieur (CINES, Montpellier, France). This work was supported by the ACI “Molécules et cibles thérapeutiques” (Molecules and medicinal targets) of the French Ministry of Research. The work of Vincent Leroux was funded by an MRT fellowship related to the ACI.

References

1. Sadowski, I.; Stone, J.C.; Pawson, T. A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Mol. Cell. Biol.* **1986**, *6*, 4396-4408.

2. Pawson, T.; Gish, G.D. SH2 and SH3 domains: from structure to function. *Cell* **1992**, *71*, 359-362.

3. Moran, M.F.; Koch, C.A.; Anderson, D.; Ellis, C.; England, L.; Martin, G.S.; Pawson, T. Src homology region 2 domains direct protein-protein interactions in signal transduction. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8622-8626.

4. Smithgall, T.E. SH2 and SH3 domains: potential targets for anti-cancer drug design. *J. Pharmacol. Toxicol. Methods* **1995**, *34*, 125-132.

5. Garbay, C.; Liu, W-Q.; Vidal, M.; Roques, B.P. Inhibitors of Ras signal transduction as anti-tumor agents. *Biochem. Pharmacol.* **2000**, *60*, 1165-1169.

6. Liu, W-Q.; Vidal, M.; Mathé, C.; Périgaud, C.; Garbay, C. Inhibition of the Ras-dependant mitogenic pathway by phosphopeptide prodrugs with antiproliferative properties. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 669-672.

7. Vidal, M.; Gigoux, V.; Garbay, C. SH2 and SH3 domains as targets for anti-proliferative agents. *Crit. Rev. Oncol. Hematol.* **2001**, *40*, 175-186.

8. Lowenstein, E.J.; Daly, R.J.; Batzer, A.G.; Li, W.; Margolis, B.; Lammers, R.; Ullrich, A.; Skolnik, E.Y.; Bar-Sagi, D.; Schlessinger, J. The SH2 and SH3 domain-containing protein Grb2 links receptor tyrosine kinases to Ras signalling. *Cell* **1992**, *70*, 432-442.

9. Maignan, S.; Guilloteau, J.P.; Fromage, N.; Arnoux, B.; Becquart, J.; Ducruix, A. Crystal structure of the mammalian Grb2 adaptor. *Science* **1995**, *268*, 291-293.

10. Chardin, P.; Cussac, D.; Maignan, S.; Ducruix, A. The Grb2 adaptor. *FEBS Lett.* **1995**, *369*, 47-51.

11. Buday, L. Membrane-targeting of signalling molecules by SH2/SH3 domain-containing adaptor proteins. *Biochem. Biophys. Acta* **1999**, *1422*, 187-204.

12. Xie, Y.; Li, K.; Hung, M.C. Tyrosine phosphorylation of Shc proteins and formation of Shc/Grb2 complex correlate to the transformation of NIH3T3 cells mediated by the point-mutation activated neu. *Oncogene* **1995**, *10*, 2409-2413.

13. Salcini, A.E.; McGlade, J.; Pelicci, G.; Nicoletti, I.; Pawson, T.; Pelicci, P.G. Formation of the Shc-Grb2 complexes is necessary to induce neoplastic transformation by overexpression of Shc proteins. *Oncogene* **1994**, *9*, 2827-2836.

14. Rozakis-Adcock, M.; Fernley, R.; Wade, J.; Pawson, T.; Bowtell, D. The SH2 and SH3 domains of mammalian Grb2 couple the EGF receptor of the Ras activator mSos1. *Nature* **1993**, *363*, 83-85.

15. Rozakis-Adcock, M.; McGlade, J.; Mbamalu, G.; Pelicci, G.; Daly, R.; Li, W.; Batzer, A.; Thomas, S.; Brugge, J.; Pelicci, M.G.; Schlessinger, J.; Pawson, T. Association of the Shc and Grb2/Sem5 SH2-containing proteins is implicated in activation of the Ras pathway by tyrosine kinases. *Nature* **1992**, *360*, 689-692.

16. Schlessinger, J. How receptor tyrosine kinases activate Ras. *Trends Biol. Sci.* **1993**, *18*, 273-275.

17. Khosravi-Far, R.; Der, C.J. The Ras signal transduction pathway. *Cancer Metastasis Rev.* **1994**, *13*, 67-89.

18. Davies, H.; Bignell, G.R.; Cox, C.; Stephens, P.; Edkins, S.; Clegg, S.; Teague, J.; Woffendin, H.; Garnett, M.J.; Bottomley, W.; Davis, N.; Dicks, E.; Ewing, R.; Floyd, Y.; Gray, K.; Hall, S.; Hawes, R.; Hugues, J.; Kosmidou, V.; Menzies, A.; Mould, C.; Parker, A.; Stevens, C.; Watt, S.; Hooper, S.; Wilson, R.; Jayatilake, H.; Gusterson, B.A.; Cooper, C.; Shipley, J.; Hargrave, D.; Pritchard-Jones, K.; Maitland, N.; Chenevix-Trench, G.; Riggins, G.J.; Bigner, D.D.; Palmieri, G.; Cossu, A.; Flanagan, A.; Nicholson, A.; Ho, J.W.; Leung, S.Y.; Yuen, S.T.; Weber, B.L.; Siegler, H.F.; Darrow, T.L.; Paterson, H.; Marais, R.; Marshall, C.J.; Wooster, R.; Stratton, M.R.; Futreal, P.A. Mutations of the BRAF gene in human cancer. *Nature* **2002**, *417*, 906-907.

19. Pendergast, A.M.; Quilliam, L.A.; Cripe, L.D.; Bassing, C.H.; Dai, Z.; Li, N.; Batzer, A.; Rabun, K.M.; Der, C.J.; Schlessinger, J. BCR-Abl-induced oncogenesis is mediated by direct interactions with the SH2 domain of the Grb2 adaptor protein. *Cell* **1993**, *75*, 175-185.

20. Daly, R.J.; Binder, M.D.; Sutherland, R.L. Overexpression of the Grb2 gene in human breast cancer cell lines. *Oncogene* **1994**, *9*, 2723-2727.

21. Boutin, J.A. Tyrosine protein kinase inhibition and cancer. *Int. J. Biochem.* **1994**, *26*, 1203-1226.

22. Brugge, J.S. New intracellular targets for therapeutic drug design. *Science* **1993**, *260*, 918-919.

23. Songyang, Z.; Shoelson, S.E.; Chaudhuri, M.; Gish, G.; Pawson, T.; Haser, W.G.; King, F.; Roberts, T.; Ratnofsky, S.; Lechleider, R.J.; Nell, B.G.; Birge, R.B.; Fajardo, J.E.; Chou, M.M.; Hanafusa, H.; Schaffhausen, B.; Cantley, L.C. SH2 domains recognize specific phosphopeptide sequences. *Cell* **1993**, *72*, 767-778.

24. Songyang, Z.; Shoelson, S.E.; McGlade, J.; Olivier, P.; Pawson, T.; Bustelo, X.R.; Barbacid, M.; Sabe, H.; Hanafusa, H.; Yi, T.; Ren, R.; Baltimore, D.; Ratnofsky, S.; Feldman, R.A.; Cantley, L.C. Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, Grb2, HCP, SHC, Syk, and Vav. *Mol. Cell. Biol.* **1994**, *14*, 2777-2785.

25. Rahuel, J.; Gay, B.; Erdmann, D.; Strauss, A.; García-Echeverría, C.; Furet, P.; Caravatti, G.; Fretz, H.; Schoepfer, J., Grütter, M.G. Structural basis for specificity of GRB2-SH2 revealed by a novel ligand binding mode. *Nat. Struct. Biol.* **1996**, *3*, 586-589.

26. Nioche, P.; Liu, W.Q.; Broutin, I.; Charbonnier, F.; Latreille, M.T.; Vidal, M.; Roques, B.; Garbay, C.; Ducruix, A. Crystal structures of the SH2 domain of Grb2: Highlight on the binding of a new high-affinity receptor. *J. Mol. Biol.* **2002**, *315*, 1167-1177.

27. García-Echeverría, C.; Furet, P.; Gay, B.; Fretz, H.; Rahuel, J.; Schoepfer, J.; Caravatti, G. Potent antagonists of the SH2 domain of Grb2: optimization of the X+1 position of 3-amino-Z-Tyr(PO_3H_2)-X+1-Asn-NH_2. *J. Med. Chem*. **1998**, *41*, 1741-1744.

28. Gay, B.; Furet, P.; García-Echeverría, C.; Rahuel, J.; Chêne, P.; Fretz, H.; Schoepfer, J.; Caravatti, G. Dual specificity of Src homology 2 domains for phosphotyrosine peptide ligands. *Biochemistry* **1997**, *36*, 5712-5718.

29. Liu, W-Q.; Vidal, M.; Gresh, N.; Roques, B.P.; Garbay, C. Small peptides containing phosphotyrosine and adjacent αMe-phosphotyrosine or its mimetics as highly potent inhibitors of Grb2 SH2 domain. *J. Med. Chem.* **1999**, *42*, 3737-3741.

30. Eck, M.J.; Shoelson, S.E.; Harrison, S.C. Recognition of a high affinity phosphotyrosyl peptide by the Src homology 2 domain of p56lck. *Nature* **1993**, *362*, 87-91.

31. Machida, K.; Mayer, B.J. The SH2 domain: versatile signaling module and pharmaceutical target. *Biochim. Biophys. Acta* **2005**, *1747*, 1-25.

32. Thornton, K.H.; Mueller, W.T; McConnell, P; Zhu, G.; Saltiel, A.R.; Thanabal, V. Nuclear magnetic resonance solution structure of the growth factor receptor-bound protein 2 Src homology 2 domain. *Biochemistry* **1996**, *35*, 11852-11864.

33. Kimber, M.S.; Nachman, J.; Cunningham, A.M.; Gish, G.D.; Pawson, T.; Pai, E.F. Structural basis for specificity switching of the Src SH2 domain. *Mol. Cell* **2000**, *5*, 1043-1049.

34. Rahuel, J.; García-Echeverría, C.; Furet, P.; Strauss, A.; Caravatti, G.; Fretz, H.; Schoepfer, J.; Gay, B. Structural basis for the high affinity of amino-aromatic SH2 phosphopeptide ligands. *J. Mol. Biol.* **1998**, *279*, 1013-1022.

35. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucl. Acids Res.* **2000**, *28*, 235-242.

36. Ettmayer, P.; France, D.; Gounarides, J.; Jarosinski, M.; Martin, M-S.; Rondeau, J-M.; Sabio, M.; Topiol, S.; Weidmann, B.; Zurini, M.; Bair, K.W. Structural and conformational requirements for high-affinity binding to the SH2 domain of Grb2. *J. Med. Chem.* **1999**, *42*, 971-980.

37. Gresh, N.; Liu, W-Q.; Garbay, N. Unpublished work.

38. Gilmer, T.; Rodriguez, M.; Jordan, S.; Crosby, R.; Alligood, K.; Green, M.; Kimery, M.; Wagner, C.; Kinder, D.; Charifson, P.; Hassell, A.M.; Willard, D.; Luther, M.; Rusnak, D.; Sternbach, D.D.; Mehrotra, M.; Peel, M.; Shampine, L.; Davis, R.; Robbins, J.; Patel, I.R.; Kassel, D.; Burkhart, W.; Moyer, M.; Bradshaw, T.; Berman, J. Peptide inhibitors of src SH3-SH2-phosphoprotein interactions. *J. Biol. Chem.* **1994**, *269*, 31711-31719.

39. Lubman, O.Y.; Waksman, G. Structural and thermodynamic basis for the interaction of the Src SH2 domain with the activated form of the PDGF β-receptor. *J. Mol. Biol.* **2003**, *328*, 655-668.

40. Henriques, D.A.; Ladbury, J.E. Inhibitors to the Src SH2 domain: A lesson in structure-thermodynamic correlation in drug design. *Arc. Biochem. Biophys.* **2001**, *390*, 158-168.

41. Hanessian, S.; McNaughton-Smith, G.; Lombart, H-G.; Lubell, W.D. Design and synthesis of conformationally constrained amino acids as versatile scaffolds and peptide mimetics. *Tetrahedron* **1997**, *53*, 12789-12854.

42. Eguchi, M.; Lee, M.S.; Nakanishi, H.; Stasiak, M.; Lovell, S.; Kahn, M. Solid-phase synthesis and structural analysis of bicyclic β-turn mimetics incorporating functionality at the I and I+3 positions. *J. Am. Chem. Soc.* **1999**, *121*, 12204-12205.

43. Eguchi, M.; Lee, M.S.; Stasiak, M.; Kahn, M. Solid-phase synthesis and solution structure of bicyclic β-turn peptidomimetics: diversity at the I position. *Tetrahedron Lett.* **2001**, *42*, 1237-1239.

44. Liu, W-Q.; Olszowy, C.; Bischoff, L.; Garbay, C. Enantioselective synthesis of (2S)-2-(4-phosphonophenylmethyl)-3-aminopropanoic acid suitably protected for peptide synthesis. *Tetrahedron Lett.* **2002**, *43*, 1417-1419.

45. Furet, P.; Gay, B.; Caravatti, G.; García-Echeverría, C.; Rahuel, J.; Schoepfer, J.; Fretz, H. Structure-based design and synthesis of high affinity tripeptide ligands of the Grb2 SH2 domain. *J.Med.Chem.* **1998**, *41*, 3442-3449.

46. Ogura, K.; Tsuchiya, S.; Terasawa, H.; Yuzawa, S.; Hatanaka, H.; Mandiyan, V.; Schlessinger, J.; Inagaki, F. Solution structure of the Sh2 domain of Grb2 complexed with the Shc derived phosphotyrosine containing peptides. *J. Mol. Biol.* **1999**, *289*, 439-445.

47. Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comp. Phys.* **1999**, *151*, 283-312.

48. Ray, N.; Cavin, X.; Paul, J-C.; Maigret, B. Intersurf: dynamic interface between proteins. *J. Mol. Graph. Modell.* **2005**, *23*, 347-354.

49. De Lucca, G.V.; Erickson-Viitanen, S.; Lam, P.Y.S. Cyclic HIV proteases inhibitors capable of displacing the active site structural water molecule. *Drug Discovery Today* **1997**, *2*, 6-18.

50. Clarke, C.; Woods, R.J.; Gluska, J.; Cooper, A.; Nutley, M.A.; Boons, G-J. Involvement of water in carbohydrate – protein binding. *J. Am. Chem. Soc.* **2001**, *123*, 12238-12247.

51. Dunitz, J.D. Win some, lose some: enthalpy – entropy compensation in weak intermolecular interactions. *Chem. Biol.* **1995**, *2*, 709-712.

52. Verdonk, M.L.; Chessari, G.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Nissink, J.W.M.; Taylor, R.D.; Taylor, R. Modelling water molecules in protein – ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504-6515.

53. Ladbury, J.E. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.* **1996**, *3*, 973-980.

54. Morton, C.J.; Ladbury, J.E. Water mediated protein – DNA interactions: The relationship of thermodynamics to structural detail. *Protein Sci.* **1996**, *5*, 2115-2118.

55. Mancera, R.L. De novo ligand design with explicit water molecules: an application to bacterial neuraminidase. *J. Comput. Aided Mol. Des.* **2002**, *16*, 479-499.

56. Tame, J.R.H.; Sleigh, S.H.; Wilkinson, A.J.; Ladbury, J.E. The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat. Struct. Biol.* **1996**, *3*, 998-1001.

57. Rostom, A.A.; Tame, J.R.H.; Ladbury, J.E.; Robinson, C.V. Specificity and interactions of the protein OppA: Partitioning solvent binding effect using mass spectrometry. *J. Mol. Biol.* **2000**, *296*, 269-279.

58. Jones, G.; Willett, P.; Glen, R.C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43-53.

59. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727-748.

60. Antony, J.; Piquemal, J-P.; Gresh, N. Complexes of thiomandelate and captopril mercaptocarboxylate inhibitors to metallo-β-lactamase by polarizable molecular mechanics. Validation on model binding sites by quantum chemistry. *J. Comp. Chem.* **2005**, *26*, 1131-1147.

61. Gresh, N.; Shi, G.B. Conformation-dependent intermolecular interaction energies of the triphosphate anion with divalent metal cations. Application to the ATP-binding site of a binuclear bacterial enzyme. A parallel quantum chemical and polarizable molecular mechanics investigation. *J. Comp. Chem*. **2004**, *25*, 160-168.

62. Langlet, J.; Claverie, P.; Caillet, J.; Pullman, A. Improvements on the continuum model. Application to the calculation of the vaporization thermodynamic quantities of nonassociated liquids. *J. Phys. Chem.* **1988**, *92*, 1617-1631

63. Phan, J.; Shi, Z-D.; Burke Jr., T.R.; Waugh, D.S. Crystal structures of a high-affinity macrocyclic peptide mimetic in complex with the Grb2 SH2 domain. *J. Mol. Biol.* **2005**, *353*, 104-115.

Multiple-step virtual screening using VSM-G.

Overview and validation of fast geometrical matching enrichment.

*Alexandre Beautrait[1], Vincent Leroux[1], Matthieu Chavent[1], Léo Ghemtio[1], Marie-Dominique Devignes[1], Malika Smaïl-Tabbone[1], Wensheng Cai[2], Xuegang Shao[2], Gilles Moreau[3], Peter Bladon[4], Jianhua Yao[5], Bernard Maigret[1]**

1 : Nancy Université, LORIA, Groupe ORPAILLEUR, Campus scientifique, BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France.
2 : Nankai University, Department of Chemistry, Tianjin 300071, P.R. China.
3 : 30 Avenue Jean Jaurès, 94220 Charanton, France.
4 : Interprobe Chemical Services, Gallowhill House, Larch Avenue, Lenzie Kirkintilloch, Glasgow G66 4HX, Scotland, UK.
5 : Shanghai Institute of Organic Chemistry, Laboratory of Computer Chemistry and Chemoinformatics, 354 Fenglin rd, Shanghai 200032, P.R. China.
* Corresponding author: bernard.maigret@loria.fr, +33 354 958 608 (telephone), +33 383 275 652 (fax).

Abstract

Numerous methods are available for use as part of a virtual screening strategy, but yet none of those is solely able to guarantee both a level of confidence comparable to experimental screening and a computing efficiency that could drastically cut down the costs of early phase drug discovery campaigns. We present here VSM-G (**V**irtual **S**creening **M**anager for computational **G**rids), a virtual screening platform that combines several structure-based drug design tools. VSM-G aims to be as user-friendly as possible while retaining enough flexibility to accommodate other *in silico* techniques as they are developed.

In order to illustrate VSM-G concepts, we present a proof-of-concept study of a fast geometrical matching method based on spherical harmonics expansions surfaces. This technique is implemented in VSM-G as the first module of a multiple-step sequence tailored for high-throughput experiments. We show that using this protocol, notable enrichment of the input molecular database can be achieved against a specific target, here the LXR nuclear receptor. The benefits, limitations and applicability of such an approach are discussed. Possible improvements of both the geometrical matching technique and its implementation within VSM-G are suggested.

Keywords

multiple-step virtual screening; VSM-G; structure-based drug design; geometrical matching; spherical harmonics surfaces; SHEF; GOLD; molecular database enrichment.

Introduction

The search for new drugs is time-consuming and expensive [1]; any method that speeds up the process is beneficial. Recently virtual screening (VS) techniques [2] have gained much interest in many drug development strategies [3]. VS has two obvious advantages: the speed with which one can screen a large library of compounds and the small initial capital investment compared to the cost of an in vitro high-throughput screening (HTS) program. The first aim of HTS and VS is to reduce a molecular database to few hit compounds for a protein target. VS is considered successful when, combined or not with HTS, it leads to confirmed hits for a cost lower to that of HTS alone. Research in this area is particularly active and several success stories have been reported [4-7]. Thus it is now widely accepted that VS calculations can complement HTS experiments [8, 9].

VS methods can have two distinct purposes. The first one is the exclusion of a large number of compounds which have little or no activity, leading to a limited set of molecules which are more probable hits [10]; such a method is referred to as a *filter*. In the literature, database filtering against a given target is often referred to as *enrichment* [11, 12]. The second purpose is the identification of a small number of candidates likely to be potent, by ranking input compounds. In all VS filters there is a trade-off between speed and accuracy; filters are optimized for speed. The fastest filters can handle up to a few million molecules, but are notoriously imprecise in reducing this number to less than a thousand while retaining all potential hits. More costly techniques, which can be used in lead optimization strategies, can tackle this problem [13, 14] but not with several million molecules as input and sensible computation times [15]. Therefore VS protocols are often based on a single or a few fast filters, and used prior to experimental screening. However, in that case, VS usage is limited to that of a pre-filter for HTS, reducing the number of compounds to be tested experimentally and hence the cost of experiments by at least one order of magnitude [16, 17].

We have devised a platform for virtual screening, called VSM-G (Virtual Screening Manager for computational Grids). Our objective with VSM-G is to provide a user-friendly tool that would give scientists a large range of *in silico* strategies for finding hits. Two kinds of approaches can be employed here – ligand-based and structure-based [18, 19]. At present VSM-G uses structure-based methods to rank input compounds according to their affinity for a target. Thus it can prioritize them for experimental testing. Ligand-based modules, such as substructure search, can be involved as pre-processing steps to screen molecular databases and reduce the number of compounds to be considered subsequently. This initial operation can precede the central element of the platform, the *screening funnel*, a multi-step structure-based filtering process that hierarchically combines several docking methods.

After describing the VSM-G platform, we will present a proof-of-concept study in the filtering/enrichment context using the liver-X-receptor β (LXRβ) as a target for a screening calculation against a diverse ligand database. The VSM-G screening funnel was used, consisting of a fast geometrical matching filter preceding flexible docking. This approach is compared to using flexible docking alone for VS. The benefits and limitations of geometrical matching as part of the screening funnel approach, in terms of computing efficiency, applicability and relevance, are discussed.

Overview of the VSM-G platform

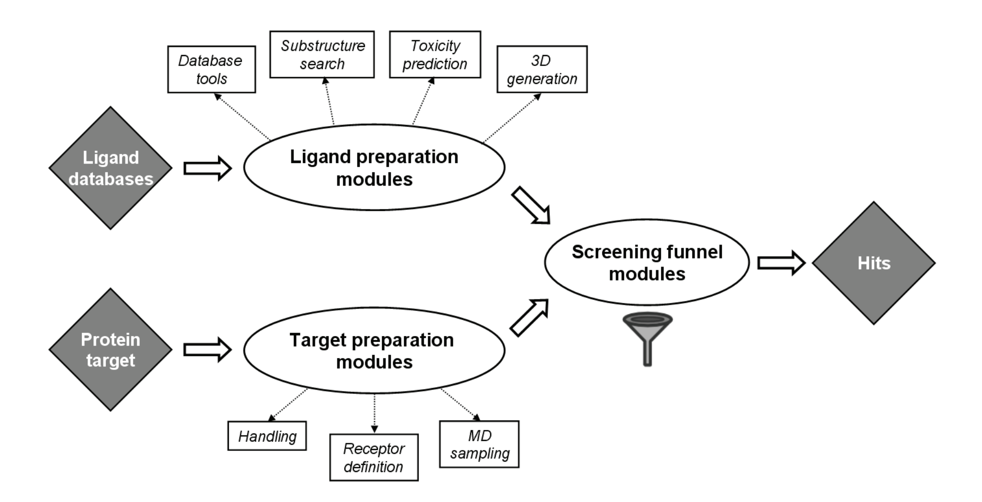
Aims and scope of VSM-G

The first step of the pre-clinical drug discovery process can be simplified as a work of exploration at the intersection of distinct spaces [20]. The first of these is the proteome, whose exploration in the drug design context involves its restriction to the sub-space of proteins whose interactions could be significant therapeutically as novel targets – the *target space*. The second space starts from the even larger ensemble of synthesizable small chemical structures. The exploration here involves sorting out molecules with no or unwanted biological effects, restraining the chemical space [21] to the so-called *drug space* [22]. Eventually, merging the target space with the drug space leads to a third ensemble of receptor-ligand associations that have to be explored successfully in order to solve the drug discovery problem. Provided that the ensembles of targets and candidate molecules have been previously reduced efficiently to avoid a combinatorial explosion, this is still a long and arduous process.

VSM-G rationalizes these searches. It focuses on the exploitation and management of current knowledge of the proteome-to-target and chemical-to-drug steps. It also relies on a specific protocol relying on structure-based virtual screening methods regarding the final ligand-to-hit process. Its workflow has been designed so to match the processes described above and is summarized in **figure 1**. The basic organization of the platform is therefore divided in three distinct parts: two for the preparation of input data (ligands and protein targets respectively) and a third one which is a multi-layer funnel for the *in silico* screening.

Figure 1

Basic VSM-G workflow for hit discovery.



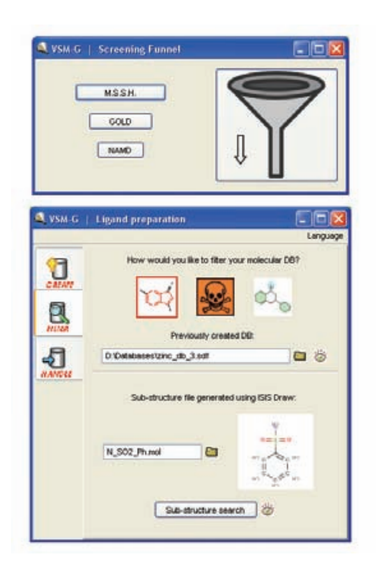
Current status

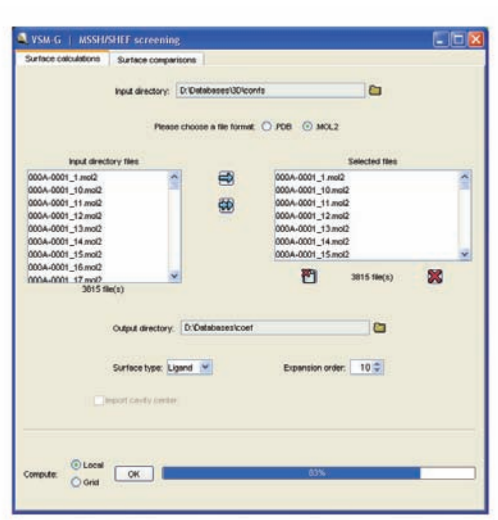
The key features of VSM-G are as follows:

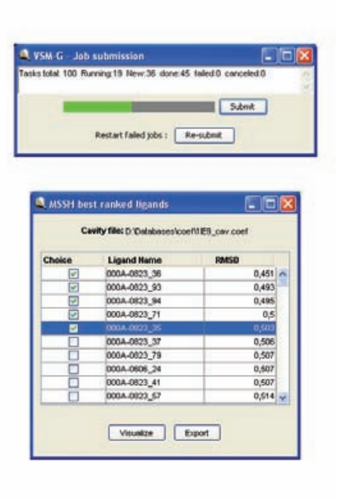
1. Wide coverage of the VS process, from ligand and target preparation to the screening setup, the monitoring of the calculation processes and finally the results' analysis.
2. Unified and user-friendly graphical interface (see **figure 2**). Seamless integration of the modules, *e.g.* intercommunication procedures, such as file format conversions, are automatic and transparent to the user.
3. Easy maintenance of the code, with modular design and choice of widely used programming languages (Java, C, C++ and Fortran).
4. Access to grid technology to take advantage of distributed computing involving computer- and cluster-grids.
5. VSM-G relies on third-party software for performing specific tasks, or in order to provide several choices of techniques for a given purpose. Due to its modular design, VSM-G is readily useable even if those external programs are not installed on the host computer. One of the main development goals of VSM-G is to provide at least one free, open-source solution for each task, which is not currently the case (*e.g.* at the moment GOLD is the only choice for performing flexible docking).

Figure 2

Some screenshots of the VSM-G graphical interface.







The VSM-G features regarding the ligand database preparation and its target-related capabilities are listed on **chart 1** and **chart 2** respectively. Current development is mostly concentrated on the screening funnel part.

Chart 1

Current VSM-G features: ligand database preparation.

Database creation and handling
- generation of virtual combinatorial libraries from chemical scaffolds and fragments
- merging of molecular files, with detection of duplicate structures
- support for different file formats, the most popular SDF [23] and MOL2 [24] as output
- conversion between formats using in-house code or OpenBabel [25]
- implementation of the MarvinBeans library [26] and VIDA [27] for database browsing (if available)

Substructure search
- flexible criteria through combinations of simple operators (and, or, not, have, at least, at most…)
- support for SMILES [28], SDF and RDF [23] as input
- internal use of a canonical topology coding that greatly reduces the complexity of the requests
- quickly searches through millions of compounds on desktop computers once the coding is performed

Toxicity prediction
- implementation of PCT [29], a carcinogenicity prediction program based on SAR
- exclusion of presumably toxic compounds
- possible enrichment of the database of substructures associated with poor chemical stability or toxicity

3D structure generation
- fragment-based 3D structure generation program
- the fragment database (> 10,000 structures) can be enhanced / extended by the user
- CORINA [30], which shares the same concept, can be used alternatively (if available)
- post-processing options: protonation (at pH = 7); conformational sampling using OMEGA [27] (if available)

Chart 2

Current VSM-G features: target preparation.

Handling of protein structures
- automatic checking and cleaning of input PDB files with respect to PDB standards [31]
- protein structures can be checked using the MOLPROBITY server [32]
- correction of protonation states: link to the H++ web server [33]
- relaxation of the hydrogen positions upon energy minimization
- link to the STING [34] web-based suite of programs for data mining

Receptor definition
- holoproteins: receptor assumed to be located at the center of mass of the ligand
- apoproteins: generation of an interactive protein 2D map with MSSH [35, 36] and VMD [37] for picking up surface receptors
- manual definition of receptors can be imported from VMD selections, and exported to funnel modules
- handling of resident water molecules, potentially useful with some docking programs [38]

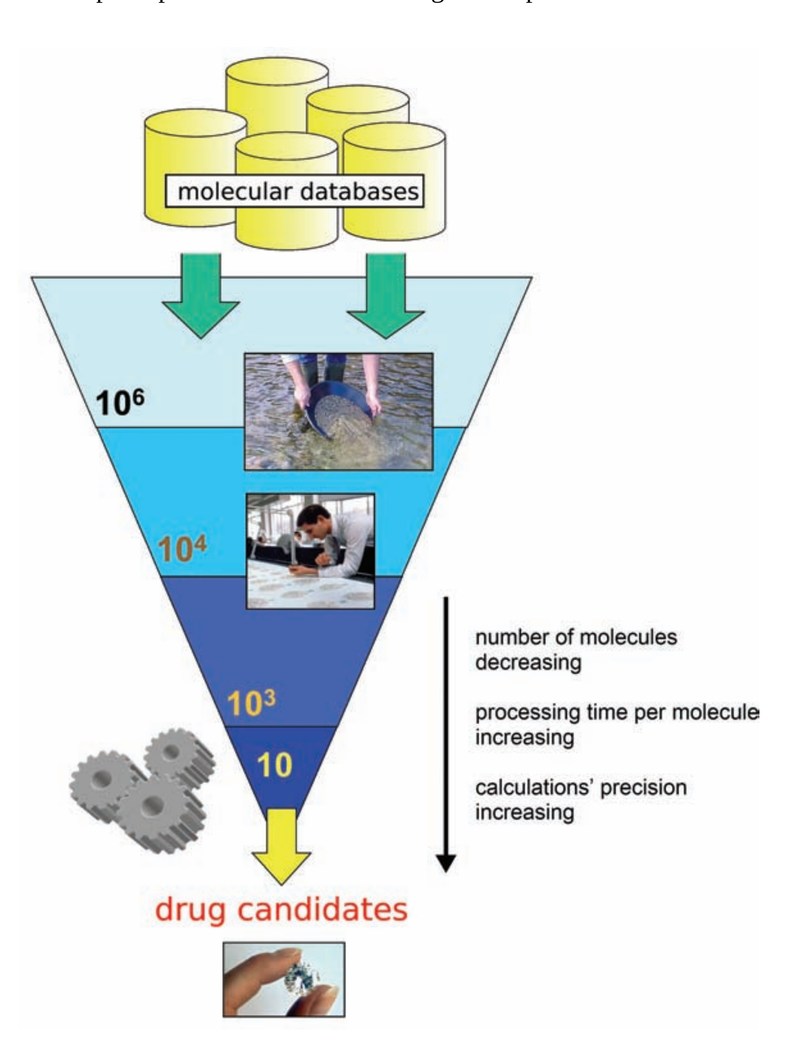
Multiple target conformations management
- handling of multiple X-ray structures
- enrichment through MD sampling [15, 39], using VMD and NAMD [40]
- clustering, averaging and minimization of conformations from NMR data or MD sampling

The screening funnel: a multiple-step strategy

Presently a wide variety of virtual screening programs are available, and it is generally assumed that a well-chosen combination of methods will give better results than a single one. The interest for such multiple-step VS protocols has been stressed in various papers, often as a combination of a single structure-based docking calculation with ligand-based approaches as pre-filters [5, 41]. Post-processing refinements starting from docking results have also been reviewed [15, 42]. Alternatively, several methods can be employed at different stages within a given docking program [43]. The use of several docking programs in the same protocol [44] is less frequent. Moreover, most programs require significant expertise in setting up and analyzing the results. More generally each technique features a specific balance between the speed of calculations and the reliability of results [45]. Open software tools overcoming such limitations are lacking. The virtual screening implementation in the VSM-G platform is constituted by a series of different structure-based methods, organized sequentially in a funnel strategy. The techniques range from simple methods to more sophisticated ones, profiting from the speed of the former and the accuracy of the latter. At each step of the process, the filter discards inappropriate compounds. The most simple and quick filters are being used at an early stage in the filtering process, allowing the more time consuming processes to be used in later stages. The multiple-step screening funnel strategy is shown in **figure 3**.

Figure 3

Basic principle of the virtual screening funnel process.



Methodology

Outline of the proof-of-concept study

Most docking methods are not efficient enough for use in high-throughput VS (*i.e.* the time required to process $>10^6$ molecules is out of reach with modern hardware). Fast filtering prior to docking might be a workaround. Ligand-based methods can also prove useful here, but unless large training sets are available for the target, they are of limited value. Geometrical matching procedures, which are orders of magnitude faster than common docking methods, can be employed in this particular context [46], and can lead to discovery of hits [47], but few studies exist to estimate their impact in a general VS experiment.

The geometrical matching procedure evaluated here is a two-step process. First, the MSSH program [35, 36] approximates the geometry of molecular structures using a series of spherical harmonics functions. This representation is very compact as all information is contained in the expansion coefficients, while the corresponding surfaces still provide a good level of detail. Additionally, this process can be done once and for all for each protein and ligand conformer. Afterwards, evaluating the surface complementarity between a target active site and a ligand is performed through simple and efficient operations [48] specific to spherical harmonics algebra. This very fast procedure is performed with the SHEF program [49], which identifies and scores the geometrically-optimal orientation of each ligand conformation for the target. These techniques are described in depth by Cai *et al.* [35, 36, 49]

In this paper we study a VSM-G-operated screening funnel using MSSH/SHEF followed by flexible docking using GOLD [50, 51]. Such an approach involves using SHEF results to filter out part of the input ligand database before proceeding to the second funnel step relying on GOLD. In this proof-of-concept study we did no such filtering; all molecules of the test set are evaluated with both techniques in order to simulate the screening funnel for all levels of filtering between the two steps.

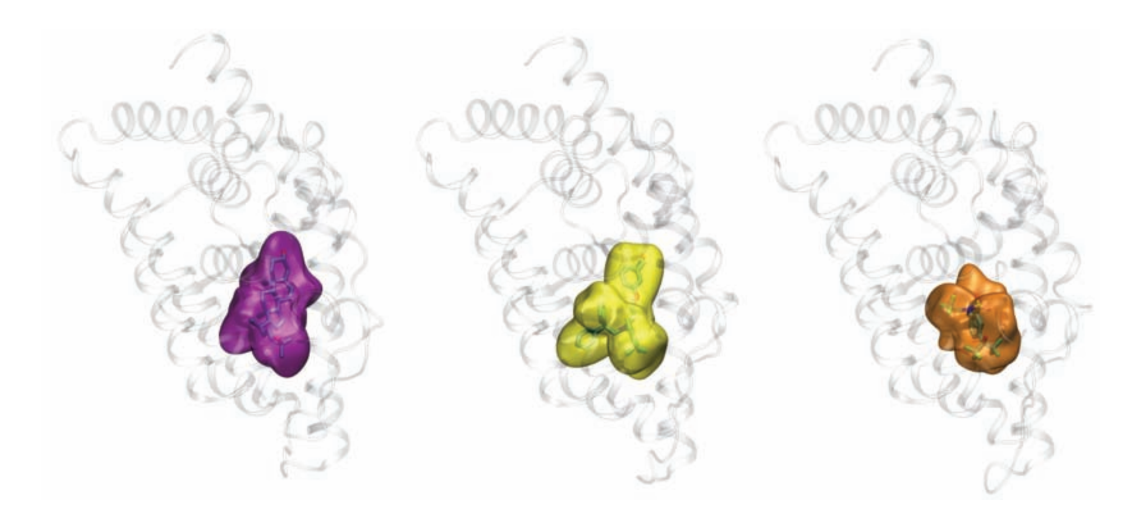
Target preparation

The liver X receptors (LXRs) [52] represent attractive targets for the development of new therapeutic agents for treating multiple (especially cardiovascular) diseases [53]. Several structures of the ligand binding domain of LXR, co-crystallized with various ligands, have been determined by X-Ray crystallography. Reports on structural analysis reveal great plasticity of the ligand binding pocket; it is able to accommodate ligands with noticeably different shapes and sizes [54]. In this work we study more specifically the LXRβ isoform, for which we took as a starting point different X-ray structures available from the Protein DataBank (PDB) [55]: 1P8D [56], 1PQ6 [54] and 1PQ9 [54]. For each of these structures the most complete chain was retained: chain A for 1P8D and chain B for 1PQ6 and 1PQ9. In all cases the binding area was complete and the Cα trace superimposed well, allowing missing fragments to be added using homology modeling. Protonation was performed at pH 7 with VSM-G. The imidazole tautomer of the active site histidine residue is the N^{d1}-H one [57].

Figure 4 shows that the three binding site conformations, represented by their MSSH-generated surfaces imported into VMD [37], are clearly distinct geometrically. The 1PQ9 cavity is significantly smaller (810 Å^3) than 1PQ6 (996 Å^3) and 1P8D (1014 Å^3). 1PQ6 has a less-spherical, more specific shape. Therefore, it could be expected that (i) 1P8D is the least selective upon ligand binding, (ii) 1PQ6 shape specificity could be overcome by ligand flexibility, and (iii) the 1PQ9 conformation should filter out more structures based on their size.

Figure 4

Shapes of the 1P8D, 1PQ6 and 1PQ9 active sites (from left to right) as approximated by spherical harmonics expansion surfaces using MSSH. The X-ray ligands filling the active sites are shown.



The protein-ligand binding modes depicted in the three experimental structures have also been analyzed. The shared characteristics are dominated by hydrophobic interactions with F_{271}, F_{329} and F_{340}. 1PQ6 allows for a possible specific charge-charge interaction with R_{319}. R_{319} already makes an internal interaction with E_{281} in the 1P8D conformation, dampening the strength of possible ligand interaction. In the case of 1PQ9 neither of those residues is accessible as the pocket size is restricted by a particular F_{329} orientation.

Ligand database preparation

The starting database is composed by compounds commercially available in March 2006 from three suppliers, ChemDiv [58], Enamine [59] and Comgenex [60]. Filtering using Lipinski's rule-of-five [61] was performed, allowing a single violation for each structure, giving a total of 598,375 unique molecules. In order to reduce the database size while retaining as many chemical diversity as possible, we used the ScreeningAssistant software [62]. This tool characterizes each molecule of the database using SSKey-3D 54-bit fingerprints [63], allowing for similarity estimation between pairs by computing Tanimoto coefficients [64]. Database clustering can then govern the generation of diversity-maximized subsets. In our case, we targeted a 10,000 molecule subset and obtained a database of 8,383 compounds.

A reference diverse database was defined by merging the initial 598,375 molecules database with the Chimiothèque Nationale (CN) database [65, 66]. Diversity of each of the three subsets (the 598,375 database, the 8,383 diversity set and the 31,220 CN) was measured as fractions of the total diversity [62]. Results are depicted in **table 1**. It appears that the 8,383 subset and the larger CN database are of comparable diversity. The former is therefore suitable as input data for a VS validation experiment. Interestingly, from the large scale database to the diversity subset we only traded ~40% of the diversity for a 98.6% size reduction.

The 8,383-compounds database was pre-processed into VSM-G ligand preparation modules, which made it suitable for the docking programs used afterwards. Molecules were first converted into 3D and then their protonation state was set arbitrarily at pH = 7. As MSSH/SHEF is a rigid shape-matching procedure, a conformational search was performed (retaining at most 400 conformers per compound), giving 1,102,299 conformers.

Table 1

Diversity analysis of the reference database used in this paper, here referred to as the *diversity subset* of 8,383 compounds. In the table, 100% diversity is that of the union of the large-scale and CN databases. All values are computed by the ScreeningAssistant software. Please refer to Monge *et al.* for details on how *drug-like* and *lead-like* compounds are defined, and how molecular database diversity is measured.

Database	number of compounds	drug-like compounds	lead-like compounds	drug-like diversity	lead-like diversity	**global diversity**
large-scale	598,327	563,777 (94.2%)	195,332 (32.6%)	84.3%	82.3%	81.8%
diversity subset	8,383	7,875 (93.9%)	3,178 (37.9%)	50.0%	43.5%	48.3%
CN	31,220	27,403 (87.8%)	20,295 (65%)	41.4%	44.8%	43.7%

Parameterization of the virtual screening programs

1,102,299 conformers were docked using SHEF in the three target conformations, giving a total of 3,306,897 rigid docking calculations. Using GOLD, 8,383 molecules were docked, giving 25,149 flexible docking calculations. The programs parameters that were used, favoring reliability over speed, are listed in **chart 3**.

Chart 3

Parameters for MSSH, SHEF and GOLD used for the validation study simulating the use of MSSH/SHEF for filtering prior to GOLD calculations.

MSSH [35, 36] **/ SHEF** [49]
- spherical harmonics expansion of order 10
- cavity coordinates defined using the ligand center of mass

GOLD [51]
- default genetic algorithm parameters
- 50 dockings / molecule
- early termination option: docking stopped if the top 5 conformations fall within 1.5 Å RMSD range
- cavity definition: flood fill (works well when the receptor is not open and extended)
- same cavity coordinates as with MSSH/SHEF
- scoring function: GoldScore

Definition and relevance of reference data

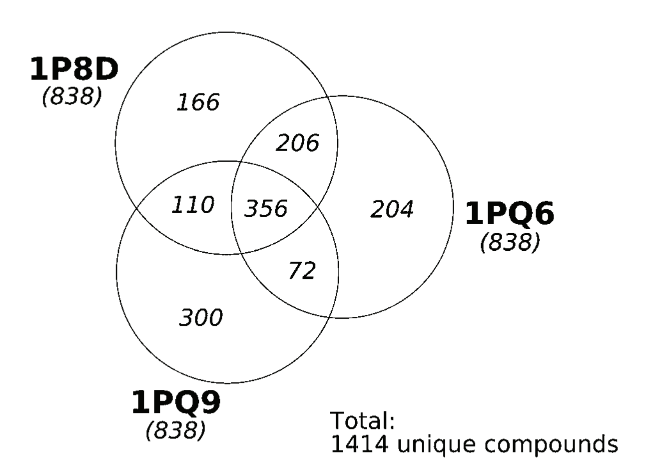
The reference data for evaluating SHEF performance is constituted by GOLD results and not by experimental data. Like all docking programs, GOLD does not provide 100% success in reproducing conformations and binding free energies of protein-ligand complexes [67]. Hence the reference set is approximate and cannot be used to measure SHEF performance precisely. However, our aim here is simply to demonstrate SHEF usefulness as part of the VSM-G screening funnel, in a large-scale VS context. Consequently, a reference set large enough statistically and chemically diverse seems appropriate despite GOLD-related limitations.

In order to evaluate filtering, the reference molecular database has to be divided in two subsets, the first corresponding to the (presumably) most potent molecules (referred to as the *hit compounds* subset) that shall be conserved upon filtering, and the second subset being considered as the inactive structures for the target. The GOLD score values are being used to rank ligands against the three target conformations, and the 10% best-ranked ligands are selected from each of the three sets. This cutoff value is set arbitrarily. Ranks are being used to select ligands instead of the score values because molecular dynamics simulations performed in our laboratory on LXRβ indicate that important induced fit effects [68] could occur upon ligand binding. This suggests the GOLD scoring function, which does not account for the receptor internal energy, may only correlate with the global free energy of binding across a single receptor conformer [69].

As shown in **figure 5**, the three ensembles of 838 selected structures overlap, giving a classification of *hits* into different families regarding their selectivity for the three target conformations. Out of a total of 1,414 molecules, 670 (47%) bind specifically on one of the three conformations, 356 (25%) bind on all the three conformations, the rest binding on two out of three. The amount of selective molecules on each conformation is 20%, 24%, and 36% for 1P8D, 1PQ6 and 1PQ9 respectively, which is in agreement with the structural specificities highlighted previously.

Figure 5

Populations of hits defined from GOLD results of the 8,383-compounds diverse database.
For each target conformation (1P8D, 1PQ6 and 1PQ9) the top-scoring 10% structures are defined as *hits*. The overlapping of these three sets is represented. There are a total of 1,414 hit compounds that is defined as the target subset that has to be conserved through the filtering process.

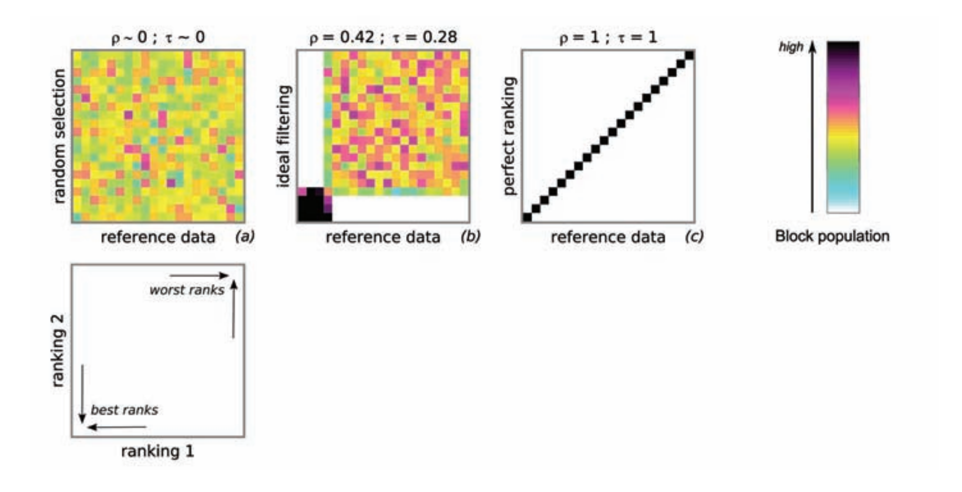


Analysis of results

An in-house program was created for representing relationships between the screening results of two different techniques for the same set of input data. **Figure 6** explains the principles of the generated graphical representation. Both ranks ranges are divided in twenty 5% blocks, a sensible trade-off between graphics clarity and the amount of represented information. Three particular cases are provided as examples. **Figure 6.a** depicts random selection, and on the opposite, **figure 6.c** corresponds to a perfect correlation. A given filtering process will obviously have results between these two. **Figure 6.b** is another ideal case for filtering, but only for a precise filtering amount (which may or may be not satisfactory).

Figure 6

Explanation of the density plots representation for ranks correlation.
Three particular cases are exemplified: (a) random selection, (b) ideal filtering, (c) perfect correlation.



The Spearman [70] ρ and Kendall [71] τ coefficients are employed as measures of correlation:

$$\rho = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^{n} \Delta r_i^{\,2}$$

$$\tau = -1 + \frac{4}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \delta(r_j > r_i)$$

r_i is the SHEF ranking of the ith-ranked GOLD structure; Δr_i is the difference between these two ranks ($\Delta r_i = r_i - i$). δ is the boolean function: δ (true) = 1 while δ (false) = 0. The rankings, in both cases, are in ascending order from the best predicted binding molecule to the worst. We also have $0 \leq \rho \leq 1$ and $-1 \leq \tau \leq 1$, 0 indicating an absence of correlation (random selection) and 1 perfect correlation (same rankings).

Other metrics are used in order to evaluate filtering performance. Given a definition of what is a hit structure and what is not for a specific target, we can describe the *quality* q of a molecular database of n structures as the ratio between the number of hit compounds and the total number of structures:

$$q = \frac{n_{hits}}{n}$$

The *enrichment* e of a database by a filtering process and for a given filtering ratio f ($0 \leq f \leq 1$, f being the amount of *filtered out* candidates) can be defined as the ratio between the quality of the reduced database and the quality of the initial database:

$$e(f) = \frac{q(f)}{q(0)}$$

Enrichment is commonly used to evaluate the efficiency of molecular database method. By definition random selection does not affect quality, so its efficiency is 1 for any filtering amount. The maximum enrichment that can be obtained for a given filtering level is when all hits are retained, which corresponds to:

$$e_{\max}(f) = \frac{1}{1-f}$$

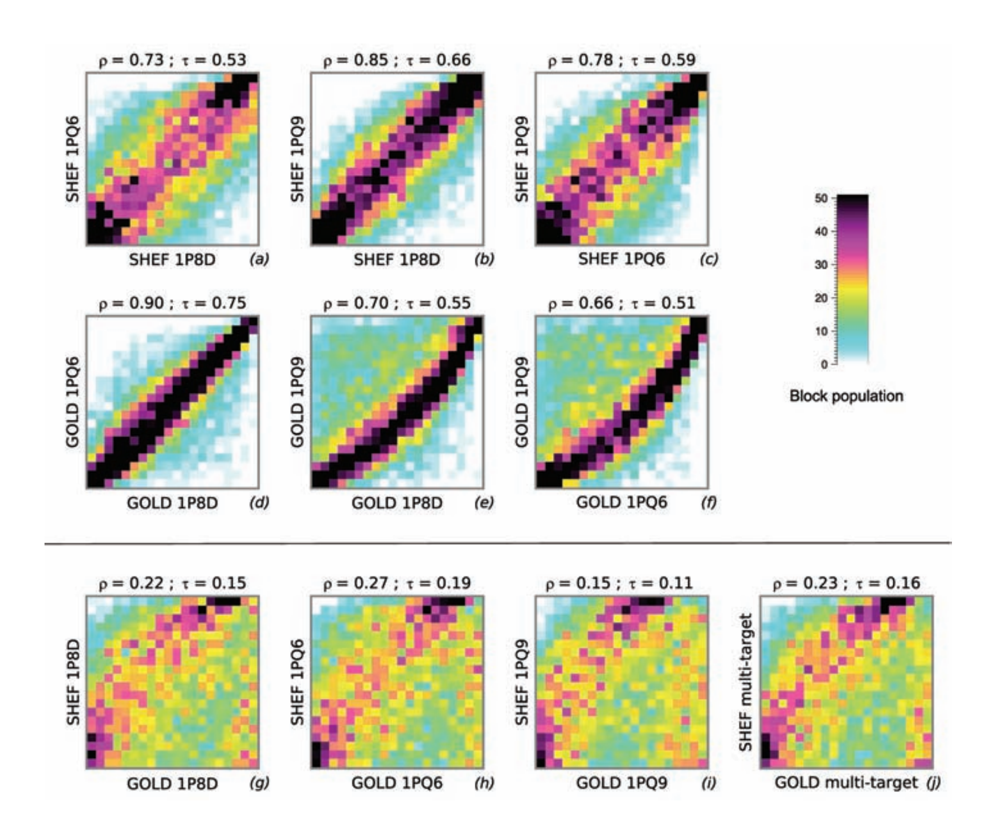
The *filtering efficiency* E is eventually defined as the relative distance of the filtering method from random filtering (E = 0) to maximum enrichment (E = 1):

$$E(f) = \frac{e(f)-1}{e_{\max}(f)-1}$$

Figure 7

Density plots between rankings.

The 6 first plots (a, b, c, d, e, f) depict the relationships between the different target conformations, for SHEF (a, b, c) and GOLD (d, e, f). Target conformation influence on these two programs can therefore be observed. The 4 last plots (g, h, i, j) show the relationship between SHEF and GOLD results, for the three target conformations (g, h, i), then using multiple-target rankings (j). The scale is set so that the average $(5\%)^2$ block density is 8383 / 400 ~ 21. Further explanations on these representations can be found on figure 6.



Results

Influence of target conformation on GOLD and SHEF results

The density plots of **figure 7** give a picture of how target conformation specificities influence GOLD and SHEF results. The SHEF correlation between 1P8D and 1PQ9 (**figure 7.b**) is greater than those between 1PQ6 and both 1P8D (**figure 7.a**) and 1PQ9 (**figure 7.c**). This is in agreement with the observation that the 1PQ6 shape is the most specific. In the case of GOLD, it first appears that 1P8D and 1PQ6 results are highly correlated (**figure 7.d**). The correlations with 1PQ9 (**figures 7.e and 7.f**) are lower. A significant amount of structures performing well with both 1P8D and 1PQ6 are ranked low with 1PQ9, indicating a group of ligands whose size fits well into the former active site conformations but not in the smaller 1PQ9. Surprisingly, such an expected group does not appear in SHEF results.

Therefore, SHEF, which is a surface-based method, appears more sensitive to the active site shape specificities than GOLD, which relies on a classical atom coordinates-based representation of molecular structures. But in contrast to GOLD, SHEF appears unable to assess size constraints correctly. This could be related not to SHEF itself but rather to its current implementation within the VSM-G screening funnel. Indeed, only the best conformer score is retained for ranking each compound: the diversity of geometrically-acceptable conformations (referred to as *adaptability*) is not taken into account. This could lead to SHEF producing false positives with ligands occupying almost all the active site volume. These ligands might require a minimal adaptability in order to provide a good chance to satisfy chemical constraints upon binding, in addition to geometrical complementarity.

Relationship between SHEF and GOLD classifications

Figures 7.g, 7.h and 7.i depict the relationships between SHEF and GOLD ranks for 1P8D, 1PQ6 and 1PQ9 respectively. Given the fundamental differences between these two programs, it is not surprising to see lower correlation between SHEF and GOLD than between two different target conformations for either SHEF or GOLD. We are, however, far from the random case depicted in **figure 6.a**, thus it is clear that noticeable enrichment using SHEF is already observed at this point.

If the general profile of the three density plots is similar, they differ regarding the distributions of false positives, *i.e.* populations located at the bottom right corners, corresponding to molecules whose binding ranks are overestimated by SHEF according to GOLD results. In agreement with previous observations, it appears that SHEF generates most false positives when docking on the 1PQ9 conformation, while correlation between GOLD and SHEF is best in the 1PQ6 case, which presents a more specific shape that should favor SHEF efficiency.

Interestingly, **figure 7.j** shows that the correlation between the SHEF and GOLD consensus rankings is higher than the average of the GOLD-SHEF correlation for the three receptor conformations. Additionally, such an approach could be more interesting than the 1PQ6-only filtering of **figure 7.h**, which naturally favors ligands more specific to 1PQ6. Even if the corresponding correlation is higher, it is probably more important to favor diversity regarding target conformations when no precise information is known concerning their relative stability.

SHEF as a first-step enrichment filter in the screening funnel protocol

It should first be noted that the ligands present in the 1P8D, 1PQ6 and 1PQ9 experimental structures, redocked using GOLD, fall into the range of the *hits* subset as defined previously. These reference ligands are also amongst the top 2% structures according to SHEF calculations. Therefore, unless the filtering ratio is set too high, they would be retrieved in a SHEF/GOLD screening funnel experiment.

Taking as reference the SHEF consensus ranking, we plotted the variation of the population of GOLD *hits* as a function of the filtering ratio. The resulting curve is shown in **figure 8** together with the enrichment curves that would result from random selection and from the ideal case where the 1,414 *hits* are all ranked before the other 6,969 molecules. A clear enrichment is observed on all ranges of filtering. There is still much room for improvement, but present SHEF performance is interesting considering that SHEF and GOLD are not in the same league in terms of speed and precision. In the virtual screening context, if the number of molecules to screen is too high for GOLD using available computing power, SHEF could provide a rational solution for decreasing the number of candidates molecules without limiting too much the chances of finding novel hit compounds for a given target.

Correlation between SHEF efficiency and the nature of the protein-ligand binding mode

We will now focus on results for two particular filtering ratios, chosen arbitrarily: 0.1 (low filtering, 90% of molecules retained) and 0.5 (half the molecules filtered out). In order to determine whether particular families of molecules could influence SHEF filtering efficiency, the variation of all *hits* populations as defined in **figure 5** for the four possible SHEF rankings (on the 1P8D, 1PQ6, 1PQ9 targets, and multiple-target consensus) was collected. The results are shown in **table 2**. This data was translated in terms of filtering efficiency E(f) in **table 3**. The main result can be interpreted as follows: if we apply respectively 10% and 50% filtering using the multiple-target SHEF filter, amongst all *hits* we will retain respectively 90.8% and 52.8% of what would have been lost using random selection.

The comparison between the four available filters based on SHEF rankings suggests that the use of the multiple-target consensus ranking should be the best choice. This is in agreement with the observations made analyzing **figures 7.g, 7.h and 7.i**. More interestingly, analysis of SHEF efficiency of the different *hits* sub-groups reveals that molecules specific to the 1PQ6 target conformation according to GOLD are performing poorly with SHEF (see **table 3**, "1PQ6-specific" line). It has been shown that the specific 1PQ6 shape is taken into account by SHEF, but 1PQ6 also presents a second particularity: the accessibility of a charged residue. The corresponding 1PQ6-specific ligands most probably share a binding mode dominated by electrostatic effects that SHEF, as it only compares geometries, is unable to assess. Contrarily, the molecules that are defined as *hits* for all of the three LXRβ pocket conformations are those for which SHEF filtering is the most efficient for both values of filtering (see **table 3**, "1P8D+1PQ6+1PQ9" line). These molecules might have a high degree of adaptability, allowing SHEF to perform well in identifying the conformations that have the best steric complementarity.

Figure 8

Enrichment curve of SHEF as measured in the validation experiment, depending on the amount of applied filtering.

Reference results are the GOLD rankings, which were used to define a target subset of 1414 structures (referred to as *hits*) out of the starting 8383. The multiple-target rankings were used in both cases (*i.e.* the rank of each molecule is the best rank amongst the 1P8D, 1PQ6 and 1PQ9 classifications). The two dotted curves represent random selection and perfect correlation (in which SHEF would reproduce GOLD results perfectly); thus the filtering efficiency E for a filtering ratio f can be measured as the relative y position of the SHEF enrichment curve between these two.

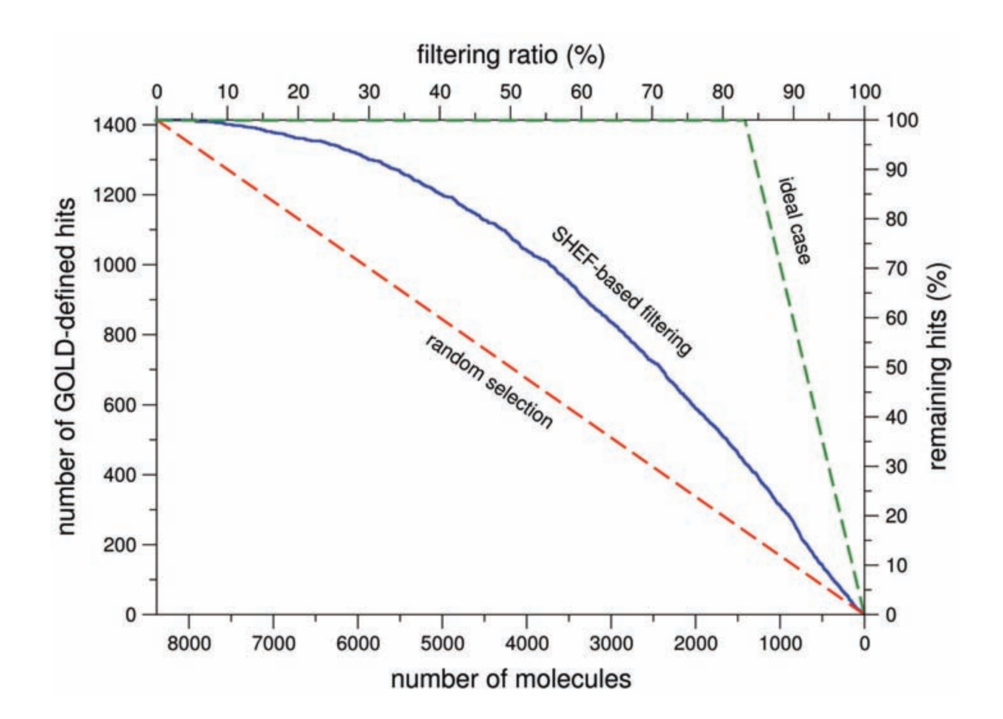


Table 2

Evolution of the GOLD *hits* subsets population when applying 10% and 50% SHEF-based filtering.

The groups defined on figure 6 are studied separately, while regarding SHEF filtering, the results for each of the 3 target conformations are presented as well as those using the multiple-target consensus ranking. Note: the multiple-target / all hits results (bottom right) can be measured directly on the figure 8 curve.

Table contents : population of the different GOLD *hit* groups after SHEF filtering		initial population	SHEF-based filters							
			1P8D		1PQ6		1PQ9		**multiple-target**	
			10%	50%	10%	50%	10%	50%	10%	50%
GOLD-based *hit* groups	1P8D	838	834	672	835	702	827	652	832	688
	1PQ6	838	828	620	833	664	817	591	826	644
	1PQ9	838	832	632	837	707	835	660	838	687
	1P8D-specific	166	165	130	165	125	164	117	165	130
	1PQ6-specific	204	198	116	201	125	192	96	197	121
	1PQ9-specific	300	295	194	299	222	297	212	300	213
	1P8D+1PQ6	206	203	156	204	151	197	137	201	142
	1P8D+1PQ9	110	110	90	110	97	110	90	110	93
	1PQ6+1PQ9	72	71	52	72	59	72	50	72	58
	1P8D+1PQ6+1PQ9	356	356	296	356	329	356	308	356	323
	all hits	1414	1398	1034	1407	1108	1388	1010	**1401**	**1080**

Table 3

Values of the SHEF filtering efficiency E(f) for f = 10% and f = 50%.
These values are directly correlated to those of table 2.

Table contents : SHEF filtering efficiency (%)		SHEF-based filters							
		1P8D		1PQ6		1PQ9		**multiple-target**	
		10%	50%	10%	50%	10%	50%	10%	50%
GOLD-based *hit* groups	1P8D	95.2	60.4	96.4	67.5	86.9	55.6	92.8	64.2
	1PQ6	88.1	48.0	94.0	58.5	74.9	41.1	85.7	53.7
	1PQ9	92.8	50.8	98.8	68.7	96.4	57.5	100	64.0
	1P8D-specific	94.0	56.6	94.0	50.6	88.0	41.0	94.0	56.6
	1PQ6-specific	70.6	13.7	85.3	22.5	41.2	-5.9	65.7	18.6
	1PQ9-specific	83.3	29.3	96.7	48.0	90.0	41.3	100	42.0
	1P8D+1PQ6	85.4	51.5	90.3	46.6	56.3	33.0	75.7	37.9
	1P8D+1PQ9	100	63.6	100	76.4	100	63.4	100	69.1
	1PQ6+1PQ9	86.1	44.4	100	63.9	100	38.9	100	61.1
	1P8D+1PQ6+1PQ9	100	66.3	100	84.8	100	73.0	100	81.5
	all hits	88.7	46.3	95.0	56.7	81.6	42.9	**90.8**	**52.8**

Discussion and concluding remarks

In this study, we wanted to present an overview of VSM-G, and then more precisely to evaluate the usefulness of the SHEF geometrical matching procedure as part of the VSM-G multiple-step high-throughput VS procedure. We have chosen, as the reference data, score values from the flexible docking program GOLD. This allows for a qualitative assessment of MSSH/SHEF efficiency as a first fast filter for the VSM-G multiple-step procedure. Thus, even considering the limitations of our validation test, results are clear enough to demonstrate that SHEF, and by extension its association as the first module in the VSM-G screening protocol, can actually be useful for *in silico* drug discovery.

This paper has highlighted precisely the conditions for obtaining good performance from MSSH/SHEF. It appears that for flexible receptors prone to induced fit effects upon complexation, a filtering based on a consensus ranking of SHEF results for multiple target conformers should be favored. More importantly, basic information regarding the types of interactions involved in ligand binding is crucial for deciding if MSSH/SHEF should be used and if so to what extent. Enrichment can only be expected when binding is not largely dominated by chemical interactions such as electrostatic effects or hydrogen bonding. Active sites that are known to favor hydrophobic interactions might be targets of choice for a structure-based drug design strategy involving MSSH/SHEF as part of a multiple-step VS procedure set up using the VSM-G program.

Limitations of the spherical harmonics-based geometrical matching procedure have been pointed out. As with all structure-based *in silico* techniques, there are two fundamental aspects of how the protein-ligand binding is modeled. Firstly, the way search space is defined, and secondly, how this space is explored. An improvement of SHEF in the first area would involve taking into account basic chemical properties to extend the complementarity score that is currently computed. Such an approach has already been tried out in the ligand-based drug design area [72]. Regarding the exploration strategy, in its current implementation in VSM-G, SHEF acts as a rigid docking program that only selects a single conformer out of a list for a given structure; this approach has been shown here to produce significant numbers of false positives in some cases. An alternative could be to use a diverse set of docked conformers for each ligand, the selection between them being made by a second module in the screening funnel protocol. Various techniques are being considered in this regard [73-75].

In any case, it is uncertain that improvements of the SHEF algorithm would necessary be worthwhile. At the present time the main advantage of the MSSH/SHEF approach is its speed. With the safe parameters used in this report, SHEF is typically 2-3 orders of magnitude faster for processing $>10^6$ conformers than GOLD for docking the corresponding $\sim10^4$ structures. MSSH is still 1 order of magnitude faster than GOLD, and its calculations can be done once and for all for a given molecular database. Enhancements of the MSSH and SHEF programs should obviously not be made at the cost of the loss of such a computing speed advantage that allows for performing large scale structure-based VS.

In further work, we will focus on selection rather than on filtering capability. This will include a proof-of-concept study of the usefulness of post-docking optimizations and molecular dynamics calculations as funnel modules following geometrical matching and flexible docking. Next, we will illustrate the whole screening funnel strategy through an actual large scale hit discovery campaign using computer grid architectures. The relevance of using advanced techniques like target sampling and grid computations in such a context will also be highlighted.

Acknowledgments

We thank Yesmine Asses, Safia Kellou and Amel Maouche for their feedback. Alexandre Beautrait was supported by grants from INRIA (Institut National de Recherche en Informatique et en Automatique), Région Lorraine, and ARC (Association pour la Recherche sur le Cancer); Vincent Leroux by a post-doctoral fellowship from the INCa (Institut National du Cancer); Matthieu Chavent by a joined fellowship between CNRS (Centre National pour la Recherche Scientifique) and Région Lorraine. We thank Openeye for providing free access to OMEGA and VIDA software according to an academic license, Chemaxon for supplying MarvinBeans Java library, CCDC for the trial version of the GOLD program, and the laboratory of chemoinformatics at the Orléans University for the ScreeningAssistant program.

References

1. DiMasi J.A., Hansen R.W. and Grabowski H.G. The price of innovation: new estimates of drug development costs. *Journal of Health Economics* **22** (2003) 151-185.

2. Shoichet B.K. Virtual screening of chemical libraries. *Nature* **432** (2004) 862-865.

3. Stahura F.L. and Bajorath J. Virtual screening methods that complement HTS. *Combinatorial Chemistry and High Throughput Screening* **7**, issue 4 (2004) 259-269.

4. Perola E., Xu K., Kollmeyer T.M., Kaufmann S.H., Prendergast F.G. and Pang Y.P. Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *Journal of Medicinal Chemistry* **43**, issue 3 (2000) 401-408.

5. Grüneberg S., Stubbs M.T. and Klebe G. Successful virtual screening for novel inhibitors of human carbonic anhydrase: Strategy and experimental confirmation. *Journal of Medicinal Chemistry* **45** (2002) 3588-3602.

6. Vangrevelinghe E., Zimmermann K., Schoepfer J., Portmann R., Fabbro D. and Furet P. Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. *Journal of Medicinal Chemistry* **46**, issue 13 (2003) 2656-2662.

7. Kraemer O., Hazemann I., Podjarny A.D. and Klebe G. Virtual screening for inhibitors of human aldose reductase. *Proteins: Structure, Function, and Bioinformatics* **55** (2004) 814-823.

8. Doman T.N., McGovern S.L., Witherbee B.J., Kasten T.P., Kurumbail R., Stallings W.C., Conolly D.T. and Shoichet B.K. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *Journal of Medicinal Chemistry* **45** (2002) 2213-2221.

9. Bajorath J. Integration of virtual and high-throughput screening. *Nature Reviews Drug Discovery* **1** (2002) 882-894.

10. Abagyan R. and Totrov M. High-throughput docking and lead generation. *Current Opinion in Chemical Biology* **5** (2001) 375-382.

11. Xu H. and Agrafiotis D.K. Retrospect and prospect of virtual screening in drug discovery. *Current Topics in Medicinal Chemistry* **2** (2002) 1305-1320.

12. Krovat E.M. and Langer T. Impact of scoring functions on enrichment in docking-based virtual screening: An application study on renin inhibitors. *Journal of Chemical Information and Computer Sciences* **44**, issue 3 (2004) 1123-1129.

13. Huo S., Wang J., Cieplak P., Kollman P.A. and Kuntz I.D. Molecular dynamics and free energy analyses of Cathepsin D-inhibitor interactions: Insight into structure-based ligand design. *Journal of Medicinal Chemistry* **45**, issue 7 (2002) 1412-1419.

14. Jenwitheesuk E. and Samudrala R. Improved prediction of HIV-1 protease inhibitor binding energies by molecular dynamics simulations. *BMC Structural Biology* **3** (2003).

15. Alonso H., Bliznyuk A.A. and Gready J.E. Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews* **26**, issue 5 (2006) 531-568.

16. Waszkowycz B., Perkins T.D.J., Sykes R.A. and Li J. Large-scale virtual screening for discovering leads in the postgenomic era. *IBM Systems Journal* **40**, issue 2 (2001) 360-376.

17. Bleicher K.H., Böhm H.-J., Müller K. and Alanine A.I. Hit and lead generation: beyond high-throughput screening. *Nature Reviews Drug Discovery* **2**, issue 5 (2003) 369-378.

18. Veselovsky A.V. and Ivanov A.S. Strategy of computer-aided drug design. *Current Drug Targets: Infectious Disorders* **3**, issue 1 (2003) 33-40.

19. Jain A.N. Virtual screening in lead discovery and optimization. *Current Opinion in Drug Discovery & Development* **7**, issue 4 (2004) 396-403.

20. Ofran Y., Punta M., Schneider R. and Rost B. Beyond annotation transfer by homology: novel protein-function prediction methods to assist drug discovery. *Drug Discovery Today* **10**, issue 21 (2005) 1475-1482.

21. Dobson C.M. Chemical space and biology. *Nature* **432** (2004) 824-828.

22. Oprea T.I. and Gottfries J. Chemography: The art of navigating in chemical space. *Journal of Combinatorial Chemistry* **3**, issue 2 (2001) 157-166.

23. http://www.mdl.com/solutions/white_papers/ctfile_formats.jsp

24. http://www.tripos.com/data/support/mol2.pdf

25. http://www.openbabel.sourceforge.net

26. http://www.chemaxon.com/products.html

27. http://www.eyesopen.com

28. Weininger D. SMILES, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* **28**, issue 1 (1988) 31-36.

29. Liao Q., Yao J.H., Li F., Yuan S.G., Doucet J.-P., Panaye A. and Fan B.T. CISOC-PCST: a predictive system for carcinogenic toxicity. *SAR and QSAR in Environmental Research* **15**, issue 3 (2004) 217-235.

30. Sadowski J. From atoms and bonds to three-dimensional atomic coordinates: Automatic model builders. *Chemical Reviews* **93** (1993) 2567-2581.

31. http://www.rcsb.org/pdb/static.do?p=file_formats/pdb/index.html

32. Davis I.W., Leaver-Fay A., Chen V.B., Block J.N., Kapral G.J., Wang X., Murray L.W., Arendall W.B., III, Snoeyink J., Richardson J.S. and Richardson D.C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, **in the press** (2007).

33. Gordon J.C., Myers J.B., Folta T., Shoja V., Heath L.S. and Onufriev A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Research* **33**, issue Web server issue (2005) W368-W371.

34. Neshich G., Mancini A.L., Yamagishi M.E., Kuser P.R., Fileto R., Pinto I.P., Palandrani J.F., Krauchenco J.N., Baudet C., Montagner A.J. and Higa R.H. STING Report: convenient web-based application for graphic and tabular presentations of protein sequence, structure and function descriptors from the STING database. *Nucleic Acids Research* **33**, issue Database issue (2005) D269-D274.

35. Cai W., Zhang M. and Maigret B. New approach for representation of molecular surface. *Journal of Computational Chemistry* **19**, issue 16 (1998) 1805-1815.

36. Cai W., Shao X. and Maigret B. Protein-ligand recognition using spherical harmonic molecular surfaces: towards a fast and efficient filter for large virtual throughput screening. *Journal of Molecular Graphics & Modelling*, issue 4 (2002) 313-328.

37. Humphrey W., Dalke A. and Schulten K. VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics* **14** (1996) 33-38.

38. Verdonk M.L., Chessari G., Cole J.C., Hartshorn M.J., Murray C.W., Nissink J.W.M., Taylor R.D. and Taylor R. Modeling water molecules in protein-ligand docking using GOLD. *Journal of Medicinal Chemistry* **48** (2005) 6504-6515.

39. Wong C.F., Kua J., Zhang Y., Straatsma T.P. and McCammon J.A. Molecular docking of balanol to dynamics snapshots of protein kinase A. *Proteins: Structure, Function, and Bioinformatics* **61**, issue 4 (2005) 850-858.

40. Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kalé L. and Schulten K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, issue 16 (2005) 1781-1802.

41. So S.-S. and Karplus M. Evaluation of designed ligands by a multiple screening method: Application to glycogen phosphorylase inhibitors constructed with a variety of approaches. *Journal of Computer-Aided Molecular Design* **15** (2001) 613-647.

42. Lyne P.D. Structure-based virtual screening: an overview. *Drug Discovery Today* **7**, issue 20 (2002) 1047-1055.

43. Wang J., Kollman P.A. and Kuntz I.D. Flexible ligand docking: A multistep strategy approach. *Proteins: Structure, Function, and Genetics* **36**, issue 1 (1999) 1-19.

44. Miteva M.A., Lee W.H., Montes M.O. and Villoutreix B.O. Fast structure-based virtual ligand screening combining FRED, DOCK, and Surflex. *Journal of Medicinal Chemistry* **48** (2005) 6012-6022.

45. Leroux V. and Maigret B. Should structure-based virtual screening techniques be used more extensively in modern drug discovery? *Computers and Applied Chemistry* **24**, issue 1 (2007) 1-10.

46. Yamagishi M.E.B., Martins N.F., Neshich G., Cai W., Shao X., Beautrait A. and Maigret B. A fast surface-matching procedure for protein-ligand docking. *Journal of Molecular Modeling* **12** (2006) 965-972.

47. Singh J., Chuaqui C.E., Boriack-Sjodin P.A., Lee W.C., Pontz T., Corbley M.J., Cheung H.-K., Arduini R.M., Mead J.N., Newman M.N., Papadatos J.L., Bowes S., Josiah S. and Ling L.E. Successful shape-based virtual screening: the discovery of a potent inhibitor of the type I TGFß receptor kinase (TßRI). *Bioorganic & Medicinal Chemistry Letters* **13**, issue 24 (2003) 4355-4359.

48. Ritchie D.W. and Kemp G.J.L. Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces. *Journal of Computational Chemistry* **20**, issue 4 (1999) 383-395.

49. Cai W., Xu J., Shao X., Leroux V., Beautrait A. and Maigret B. SHEF: a vHTS geometrical filter using coefficients of spherical harmonics molecular surfaces. *Journal of Molecular Modeling*, **to be submitted** (2007).

50. Jones G., Willett P. and Glen R.C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology* **245**, issue 1 (1995) 43-43.

51. Jones G., Willett P., Glen R.C., Leach A.R. and Taylor R. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* **267** (1997) 727-748.

52. Lala D.S. The liver X receptors. *Current Opinion in Investigational Drugs* **6**, issue 9 (2005) 934-943.

53. Collins J.L. Therapeutic opportunities for liver X receptor modulators. *Current Opinion in Drug Discovery & Development* **7**, issue 5 (2004) 692-702.

54. Färnegårdh M., Bonn T., Sun S., Ljunggren J., Ahola H., Wilhelmsson A., Gustafsson J.-Å. and Carlquist M. The three-dimensional structure of the liver X receptor ß reveals a flexible ligand-binding pocket that can accommodate fundamentally different ligands. *Journal of Biological Chemistry* **278**, issue 40 (2003) 38821-38828.

55. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. and Bourne P.E. The Protein Data Bank. *Nucleic Acids Research* **28**, issue 1 (2000) 235-242.

56. Williams S., Bledsoe R.K., Collins J.L., Boggs S., Lambert M.H., Miller A.B., Moore J., McKee D.D., Moore L., Nichols J., Parks D., Watson M., Wisely B. and Willson T.M. X-ray crystal structure of the liver X receptor beta ligand binding domain: regulation by a histidine-tryptophan switch. *Journal of Biological Chemistry* **278**, issue 29 (2003) 27138-27143.

57. Steiner T. and Koellner G. Coexistence of both histidine tautomers in the solid state and stabilisation of the unfavourable Nδ-H form by intramolecular hydrogen bonding:crystalling L-His-Gly hemihydrate. *Chemical Communications (Cambridge, United Kingdom)* **13** (1997) 1207-1208.

58. http://www.chemdiv.com

59. http://www.enamine.net

60. http://www.amridirect.com

61. Lipinski C.A., Lombardo F., Dominy B.W. and Feeney P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **23**, issue 1 (1997) 3-25.

62. Monge A., Arrault A., Marot C. and Morin-Allory L. Managing, profiling and analyzing a library of 2.6 million compounds gathered from 32 chemical providers. *Molecular Diversity* **10**, issue 3 (2006) 389-403.

63. Xue L., Godden J. and Bajorath J. Database searching for compounds with similar biological activity using short binary bit string representations of molecules. *Journal of Chemical Information and Computer Sciences* **39**, issue 5 (1999) 881-886.

64. Tanimoto T.T. Non-linear model for a computer assisted medical diagnostic procedure. *Transactions of the New York Academy of Sciences* **2**, issue 23 (1961) 576-580.

65. Hibert M. and Haiech J. Des gènes aux médicaments : nouveaux défis, nouvelles stratégies. *M/S : Médecine Sciences* **16**, issue 12 (2000) 1332-1339.

66. http://chimiotheque-nationale.enscm.fr/

67. http://www.ccdc.cam.ac.uk/products/life_sciences/validate/gold_validation/

68. Koshland Jr. D. The key–lock theory and the induced fit theory. *Angewandte Chemie, International Edition in English* **33**, issue 23-24 (1994) 2375-2378.

69. Redocking experiments of LXRß reference ligands present in the X-ray structures back up this hypothesis. Using GOLD, the 1PQ6 ligand redocked in the 1PQ6 binding pocket conformation yields a significantly higher score than the 1PQ9 ligand redocked in the 1PQ9 conformation. However, according to experimental data, the 1PQ9 ligand is indeed clearly more potent on LXRß than the 1PQ6 one, further indicating that the protein-ligand interaction could not be the dominant term in the free energy of binding.

70. Spearman C. The proof and measurement of associtaion between two things. *American Journal of Psychology* **15**, issue 1 (1904) 72-101.

71. Kendall M. A new measure of rank correlation. *Biometrika* **30**, issue 1-2 (1938) 81-89.

72. Mavridis L., Hudson B.D. and Ritchie D.W. Toward high throughput 3D virtual screening using spherical harmonic molecular surface representations. *Journal of Chemical Information and Modeling* **47** (2007) 1787-1796.

73. Massova I. and Kollman P.A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspectives in Drug Discovery and Design* **18**, issue 1 (2000) 113-135.

74. Gilson M.K. and Zhou H.-X. Calculation of protein-ligand binding affinities. *Annual Review of Biophysics and Biomolecular Structure* **36** (2007) 21-42.

75. Marcou G. and Rognan D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *Journal of Chemical Information and Modeling* **47**, issue 1 (2007) 195-207.

Should structure-based virtual screening techniques be used more extensively in modern drug discovery?

V. Leroux and B.Maigret

Nancy université, Université H. Poincaré – Nancy I
UMR CNRS / UHP 7565, eDAM group
BP 239, 54506 Vandœuvre-les-Nancy Cedex, France

Abstract

The drug discovery processes used by academic and industrial scientists are nowadays being questioned. The approaches of the pharmaceutical industry that were successful 20 years ago are simply not suitable anymore for the increasing complexity of available biological targets and the raising standards for medical safety. While the current scientific context resulting from significant developments in genomic, proteomic, organic synthesis and biochemistry seems particularly favorable, the efficiency of drug research does not appear to be following the trend. In particular, the in silico approaches, often considered as potential enhancements for classic drug discovery, are an interesting case. Techniques such as virtual screening did undergo many significant progresses in the past 5-10 years and have proven their usefulness in hit discovery approaches for who wants to avoid carrying out too many expensive experimental tests while exploring an important molecular diversity. However, reliability is still deceiving despite constant enhancements, and results are unpredictable. What are the origins of such issues?

In this short review, we will first summarize the current status of computer-aided drug design, then we will focus on the structure-based class of virtual screening approaches, for which docking programs constitute the main part. Can such methods give something more than cost savings in the early banks-to-hit phases of the drug discovery process? We will try to answer this question by exploring the highlights and pitfalls of the great variety of docking approaches. It will appear that while the structure-based drug design field is not yet ready to fulfill all of its early promises, it should still be investigated extensively and used with caution. Most interestingly, structure-based methods are best used when combined with other complementary drug design approaches such as the ligand-based ones. In this regard, they will have an increasing role to play in modern drug discovery, which is more and more interdisciplinary.

Keywords

modern drug discovery; structure-based drug design; computer-aided drug design; virtual screening; docking; hit discovery; hit-to-lead; grid computing.

Abbreviations

CADD: computer-aided drug design; VS : virtual screening; SBDD: structure-based drug design; LBDD: ligand-based drug design; HTS: (experimental) high-throughput screening; vHTS: virtual high-throughput screening; GA: genetic algorithm; QSAR: quantitative structure-activity relationships; MM: molecular mechanics; FF: (molecular mechanics) forcefield; MC: Monte-Carlo methods.

Trends in modern drug discovery

Nowadays the drug discovery process is being more and more complex and costly. As an interdisciplinary field by essence [1], it is correlated to the increasing knowledge provided by the large number of the “omics” issues and has to follow the advent of many new chemical, biological and computer techniques. The highest impact was certainly provided by the progress in computational capabilities which are now a key factor in modern drug design. [2, 3] In addition to such advances, there is an explosion of available data and knowledge that can be used to designate new targets of interest for drug research. In particular, the great efforts in genetics provided a much deeper understanding of biological processes and allowed the identification of more and more targets of medical interest. However, in that case the early enthusiasm was dampened as it appeared that biological function is indeed related mostly to interactions between biomolecules, mainly proteins.

In that regard, the study of putative targets can be conducted at different levels as, in fine, their three-dimensional structure must be taken into account to understand or predict their behavior, giving a much larger space to be investigated than the genomics sequential space itself. [4] Currently, the human genome is decrypted and the number of available protein experimental structures is growing exponentially. [5, 6] Simultaneously, as biologists increased their understanding of cell components, it became clear that simply obtaining a full list of them will not tell us how a cell works. Rather, even for a substructure that has been well characterized, there are significant difficulties in understanding how components interact to produce the observed behavior. Consequently, a level of perception higher than genomics has to be assessed, encompassing the list and the nature of the interactions between biomolecules. Even if it is an incredibly complex task, it is now admitted that the effort in the drug discovery process must be shifted towards the proteome [7] and ultimately to the interactome.

Given all those advances and considering the rate at which progresses are made more generally in all the domains associated to drug design [1], computing technology [8], and biochemical information access and management [9-11], it could be postulated that, in consequence, drug discovery would benefit from that trend and follow it. A quick look in this direction indicates, however, that this is not the case: the average cost ($800 millions in 2000) and duration (15 years) for the development of a new drug actually increases [12], as well as the failure rate amongst compounds submitted to clinical trials by the pharmaceutical industry [13]. Several popular explanations involve parameters external to the scientific world, particularly management concerns (the industry preferring to favor marketing over research) and the raising standards in medical safety. However, most probably limitations of research techniques should also be taken into account carefully. Drug research methods can be divided into two non-exclusive classes of approaches: the well-established classic experimental drug discovery [14] and the ever-growing ensemble of theoretical and computational methods, now integral parts of today’s drug design, following the most significant advances and emerging concepts of modern scientific research [15, 16].

A crucial paradigm of modern drug discovery is that most of the interesting targets that were reachable by the protocols successful 20-30 years ago have probably been investigated. The corresponding marketed drugs encompass a very limited part of the interactome (~500 biomolecules). [1] Studying other targets that could provide the basis for discovering innovative compounds (e.g. anti-cancer specific [17]), at the genomics and proteomics level, is significantly more demanding and requires the use of computer-aided techniques. [18] Data mining and visualisation techniques [19], associated with statistical analysis approaches, are without any doubt mandatory here. Molecular modelling methods [20] are also very important [21], as they directly simulate the possible relationships between molecular structure and biological function, which is the central point in proteomics.

The role of docking in computer-aided drug design

Computer-aided drug design (CADD) techniques can be broadly classified by defining two categories [22]: structure-based drug design (SBDD) [23, 24] and ligand-based drug design (LBDD) [25]. When studying the interaction of compounds with a given target of biological interest, the prerequisite for SBDD is the three-dimensional structure of the biomolecular target (SBDD is thus often referred to target-based or receptor-based drug design). For screening purposes, docking [26-30] is by far the most popular SBDD method. When this information is not available, one can resort to homology modelling [31] if enough structural data is known from analogues of the target. If this is not possible, LBDD techniques, focussed on the ligands only, often constitute an alternative choice. [25] The most known method used in that area is QSAR [32, 33], which describes molecules through a set of physico-chemical descriptors (generally related to the ligand structure) from which statistical analysis could show relationships with biological activity, and ultimately model a reliable empirical formula. It can be noted that both docking and QSAR often constitute computational challenges, as their search space (geometrical degrees of freedom or descriptors) is very large. For that reasons complex algorithms are often used, e.g. genetic algorithms (GA) [34] in the case of docking and neural networks [35, 36] for QSAR. It should also be remarked that such a distinction between SBDD and LBDD can be misleading because it refers more to the nature of the main input data (is a structure of the biological target of interest already available?) rather than to what kind of data will be manipulated. While we did use the de facto definitions here, this would benefit from some clarification. Additionally, we will see that such a frontier between SBDD and LBDD should be outshined.

In this short review, we will now focus on SBDD and molecular docking. The basic aim of docking, as a SBDD technique, is to solve the question of how a specific molecule would interact with a given biological target whose molecular structure is known. One of its current goals is to simulate HTS experiments reliably (and quickly enough) through virtual experiments (vHTS), making the virtual screening approach (VS) [23, 37-39] a potent cost-saving alternative or complement to screening campaigns. [40] Another highlight of docking is that valuable structural information is obtained, that can be used readily as a starting point for optimization using a great variety of molecular modeling techniques.

Issues and concerns about docking

It is often said that SBDD methods have proven their value, having a more than 10 years track record of successes [41-51]. But more generally, most methods of modern drug discovery are being questioned; even the efficiency of HTS techniques – one of the most important tools of pharmaceutical research – is being debated [52-55]. Regarding SBDD, even if VS techniques are now common in the pharma world, a patented drug entirely designed in silico has not been reported yet. [3] VS is also well known for its unpredictability, which is annoying for such an approach that, at first sight, seems to provide an escape route from the empirical practices in medicinal chemistry. [39] This also contradicts the “dream” of a modelling approach that could totally replace experiments, that was foreseen in the 1970s as SBDD was coined. [56] Docking programs, in particular, still boost poor accuracy, as redocking validation simulations, performed on a limited number of protein-ligand structures extracted from the PDB [5, 57], show that the current state-of-the-art solutions only provide success rates of 60-70% [58-60]. This clearly indicates that docking methods, regardless the progress that is made constantly in this area, still need to be improved.

Before attempting to improve docking methods, one has to explain why current more and more sophisticated approaches still do not perform as well as one can expect. Solving this problem is a prerequisite for answering the question raised here: should we use VS through SBDD methods more extensively? Recently, Shoichet postulated that there exist three main problems that need to be assessed [39] : (1) the possible chemical space is so wide that most probably we are not able to explore all of its possibilities, (2) the structure of biomolecules is very complex, particularly regarding flexibility and induced fit effects, and (3)

the efficient simulation of ligands binding capabilities, related to free energy calculations, is a difficult task. In another interesting opinion, Kubinyi indicates that problems could rather originate mostly from oversimplifications and wrong assumptions made by scientists performing in silico research, relying too much on computer results and forgetting that proper knowledge and expertise (the ”in cerebro research”) is always the first requirement. [61] While there is a large consensus around the directions indicated by Shoichet and others that motive most of the propositions for enhancing current docking methods [62], experience made us to focus more on the importance of some issues raised by Kubinyi. Docking programs, as well as SBDD techniques, are very likely to give the inexperienced user a false sense of security: by controlling and visualizing directly what ”happens” on the screen in structural details it is easy to assume that there is no problem as long as there is no sign of the contrary. This can be a fatal mistake as, above all, molecular modelling requires rigor and prudence. More specifically, setting up a docking simulation without carefully checking how the internals of the program work and how this could be related to the results, is often a straight road to more or less ”strange” failures.

The field of docking is constantly growing with new programs being introduced regularly, fuelling a strong competition. In most cases, the programs are being benchmarked for their ability to reproduce known structures, while less often the ranking capability on a reference VS experiment, including more or less potent binders as well as decoys, is also tested [58]. Emphasis is put on reliability, which is of course of premium importance for who wants to pick the most suitable program for his needs. However, in our opinion, such information is just not enough: relying solely on benchmark results eludes the fact that performing docking requires control over the process. We will now survey a representative set of docking programs focusing on these aspects.

How do docking programs work?

The choice of docking methods is large (see table I for providing a limited set of all docking programs) and evaluating them is difficult for at least four reasons: (1) each method corresponds to at least one ratio between speed and value of results (two methods that clearly do not play in the same league here most probably do not have the same applicability in VS), (2) the efficiency of a given method is more or less system-dependant, (3) each program is specific regarding its requirements concerning the nature and format of input data, parameters and results (establishing a benchmark in such a context could be very tedious), (4) most importantly, each method has its very own way to model the system, define the search space and evaluate conformations.

In this regard, rather than focusing on precision and speed, we defined the main internals of docking programs, presented in decreasing importance order. Firstly, decisions have to be taken regarding molecules and target flexibility. Typically, if the docking simulations are performed on the 3D structure of a target extracted from an X-ray structure of this target associated with a given ligand, then if the receptor flexibility treatment is too limited, the docking program will artificially favor the molecules that share the binding mode corresponding to the ligand present in the X-ray complex. If other binding modes can exist because of large induced fit effects, the docking is biased. The scoring function and search engine are the driving forces of the docking algorithms and are therefore directly related to their efficiency. For example, the GOLD docking program, having a very good search engine and a relatively weak scoring function, is good at predicting structures but is known to perform poorly if the score value is used to rank VS results. Next, solvation handling appears mandatory in some systems, as discrete water molecules can bridge specific protein-ligand interactions. [63, 64] Not permitting the treatment of such waters would again introduce a harmful bias and the programs are not equal in that area. Target definition must also be taken into account if it is not detailed precisely prior to docking. Some programs can construct (more or less efficiently) the active site if they are given a single starting coordinate, while others can try to find it when it is not known. In some cases the lack of information can simply prevent docking. The nature of the user interface and the possibility of launching docking jobs in parallel make programs more or less convenient to use, but in the case of vHTS,

the absence of parallel execution can render calculation execution and management particularly painful. Lastly, the legal status is in our opinion quite important. Not being able to inspect the program code means that the docking program is a sealed ”black box”. The users have to delegate part of the control they could exert on their simulations, provided they have the technical skills to track bugs or customize an external source code.

The information presented on the accompanying table raises a serious consideration. Amongst the programs that were presented in Table I (DOCK, GOLD, AutoDock, FlexX, ICM-Dock and Glide being the most popular, while SHEF, DARWIN and MORDOR were added for diversity) there is not a single consensus on one of the key features presented. This is not very trusty at first sight, as docking foundations appear foggy to say the least. But as a consequence of such a total heterogeneity, highlights and weaknesses are quite distinct and method-specific. A well-chosen combination of techniques (e.g. fast and crude to slow and precise) could lead to the limitations of each method being compensated by features of the others and programs being judged with regard to complementarities rather than to individual performance. Such an approach, termed consensus scoring [65, 66] is already attempted regarding scoring functions alone. Its extension to docking techniques is the basis of Glide’s ”hierarchical docking” over which unfortunately the user do not have total control, Glide being a commercial program. Ultimately, such an approach should be extended outside the boundaries of SBDD. This is the deep-seated motive of the VSM-G project. [67]

The drug discovery process: can docking and structure-based design be integrated more efficiently?

The drug discovery process which includes CADD can be simplified into several well-defined steps. One can first start from the chemical space which includes all molecules that a chemist could make. The size of this galaxy is crudely estimated to 1060. [68] Of course, we cannot explore that either experimentally or by using computing. What we need is to focus on the biochemical space – the ensemble of molecules that are relevant to biology. For that purpose, we can use pharmacokinetics [69] and various other approaches [70, 71] to estimate a drug-like probability of molecules. Simple empirical guidelines are proposed, such as the well-known Lipinski’s rule-of-five. [72] Molecular databases merging compounds from suppliers and applying basic such drug-like filtering are currently available. [73, 74]

The next steps of drug discovery aim to further reduce the numbers of molecules considered, from ~108 drug-like candidates to ~103 hits, then a limited number of leads which after optimization and careful inspection could be proposed for clinical trials. We define hits as molecules confirmed active regarding the biological target in vitro, while leads are hits which are also active in vivo. An excellent review of the associated protocols, including definitions, strategies and issues is available. [75] At both the drug candidates-to-hit (VS) and hit-to-lead (optimization) steps experiments can be conducted. In the former case HTS can greatly benefit from its association with vHTS [40, 76], but the complexity and heterogeneity of docking methods should be taken into account.

SBDD is not limited to docking techniques. Outside the VS approach, when hits have to be optimized to the largest extent possible, more versatile and much more computationally costly methods such as molecular dynamics [77-79] and free energy calculations [80] are clearly more suited. It is worthy to note that contrary to docking programs depicted on Table I, the internals of the most popular molecular dynamics packages [81-93] share the same theoretical principles, differences between the forcefields and the integrators being more technical than conceptual. Furthermore, unlike vHTS that can cut down costs drastically in hit discovery, computer-aided lead discovery techniques add more cost to the hit-to-lead phase. As the admission charge for candidates to clinical trials is so high, rushing through experiments is absolutely out of the question at this point. Moreover, no in silico technique can at present handle realistically the degree of specificity of a given interaction, even if there are early attempts in that direction. [94]. Consequently experimental approaches of classic biochemistry will continue to dominate the lead discovery phase. This does not mean that SBDD is a waste of time in the hit-to-lead phase, as a great deal of meaningful

knowledge, sometimes unreachable experimentally (e.g. local structural dynamics cannot be observed without resorting to molecular dynamics), can be potentially gained. When such data is reinserted, it is likely to enrich the whole drug discovery process. Regarding these points, Figure 1 highlights how CADD could be of use in modern drug discovery protocols.

To conclude this part, we will say that concerns about docking and SBDD being inadequately integrated in drug discovery could appear linked to misunderstandings regarding the abilities of the different methods. In SBDD the hit discovery (screening) and lead discovery (predicting) techniques are clearly distinguished, while in LBDD this is more likely only a matter of size and pertinence of data. It should be noted that some recent developments may start to fill the methodological gap between automated screening and manual modeling. [95, 96] More importantly, as stated before, docking limitations are mostly conceptual and protocols making wise use of the complementarities between different techniques could boost the efficiency of VS simulations considerably.

Perspectives for structure-based drug design

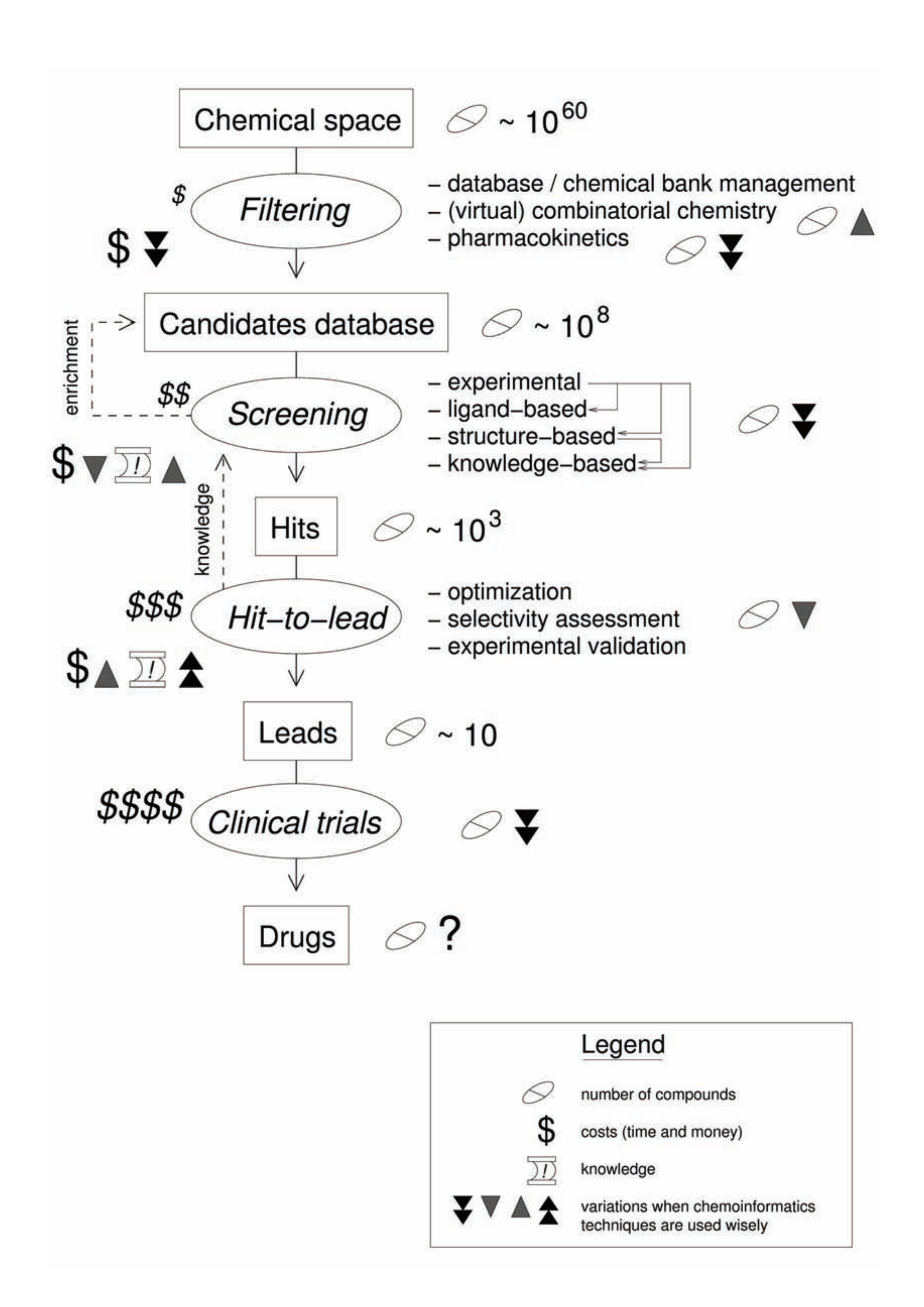
As purely empirical LBDD methods are correlated from the start to experimental data, validation is part of the design. On the opposite, SBDD methods rely on molecular models whose design depend on existing concepts, algorithmically constraints and assumptions at different levels: the validation is all but obvious and can ultimately appear to invalidate design. Considering this, it is striking to notice that, in the field of molecular docking, the development of new methods is observed at a very fast pace while the validation of existing ones is still prone to improvements [61]. Unlike LBDD closely bound to applied biochemistry, with SBDD we are often on the verge of theoretical chemistry. Taming SBDD is therefore difficult by nature, despite some appearances to the contrary. SBDD also needs to mature.

On the other hand, it can be postulated that the great variety of SBDD proposals is a chance for who possesses enough expertise. Constructing protocols using complementary docking techniques as a solution to the VS accuracy issues is an approach that should be extended to the other CADD techniques as well as to advances in data mining and computing science. Interfaces overcoming the obstacles associated with heterogeneous programs, building bridges between LBDD, SBDD, database filtering techniques, and more generally knowledge in drug design, are strongly desired. One of the goals of the post-genomic era is indeed to benefit fully from the explosion of biological, chemical and structural data, for which web-based services [9-11, 57, 74, 97] and specific databases [5, 57, 73, 74, 97-103] provide unprecedented access. Its application to drug discovery certainly involves linking this data to the variety of methods of SBDD, that will most probably benefit more and more from the advances in modern computing, such as peer-to-peer networks [15, 104-107] and computational grids [108-111].

Interestingly, the knowledge enrichment is currently taking off inside the boundaries of specific approaches. The gathering of SAR data, starting to link LBDD with modern biological knowledge, has been reported recently. [112] The VSM-G platform currently being developed [67] at the SBDD VS level within the framework of computational grids, has the long-term purpose to implement a similar ”united we stand” philosophy. The fusion of the great richness of complementary approaches, benefiting from the explosion of genomics and proteomics data, should eventually lead in the future to a more efficient global CADD driving concept that we could name knowledge-based drug design.

In conclusion, to our question regarding if SBDD VS techniques should be used more extensively, the answer is yes, but we have to specify that the focus should be on the association with the other techniques of modern drug discovery. Resulting views should be more and more linked with the advances in understanding the mechanisms of life, from the genome to the interactome. This open approach is promising, while on the exclusive side, inventing new docking techniques could be necessary for improving the tools, but as the choice there is somewhat cluttered and presently limited to hit discovery, this should not be overemphasized.

Figure 1: flowchart of the computer-aided drug discovery process



Program name	Ligand flexibility	Receptor flexibility	Scoring function	Search engine	Solvation handling	Receptor definition requirements	User interface	Parallelized?	Legal status	Highlights
DOCK [26, 113, 114]	Anchor-first strategy: ligand only flexible when optimized (minimization).	Limited rearrangements of receptor structure are possible during optimization, if AMBER score is used.	Several possible. AMBER score based mostly on the FF intramolecular terms (vdW + Coulomb). Ligand rigid segments used for the internal lig/lig term.	Geometrical grid matching + local minimization, or incremental construction + random search. [115]	Various implicit approaches depending of scoring function used.	Grid generated from center point or defined manually.	Command-line with parameter file.	Yes: MPI [116, 117].	Free to academics, source code provided.	Very versatile, particularly active support.
GOLD [118-120]	Full.	Very limited: possible rotation of some amino acids' terminal groups.	2 crude empirical scoring functions. Use of an external function possible but requires programming with supplied GOLD APIs.	GA [34].	Implicit (scoring function). Switching of explicitly modeled waters, if any [121].	Constructed from reference point/atom (risky with "open shape" receptors), from the conformation of a bound reference ligand, or specified explicitly.	Parametrization through a single configuration file. Jobs started either from the shell or the GUI.	Yes: PVM [122, 123].	Commercial.	Reliable at predicting bound ligand structures, easy to use.
AutoDock [124, 125]	Full.	Optional: flexible side chains and/or specified groups. A single grid that represents several receptor structures can also be used. [126]	Free-energy estimation function (MM part is a type 1 FF, and all parameters are empirical) including desolvation [127].	Several available: MC simulated annealing, GA, Lamarckian GA [127] (hybrid local search GA).	Implicit (scoring function). Structural waters can also be accounted for. [126]	Separate program autogrid, assuming center of active site at position (0; 0; 0), must be run prior to docking.	Command-line / parameter file. An optional GUI is available (AutoDockTools).	No. Distribution of jobs possible through scripts.	Free to academics, source code provided.	Combines efficient scoring and searching schemes.
FlexX [128, 129]	Partial, same as DOCK (optimization stage).	The receptor grid can represent an ensemble of target structures (FlexE module). [130]	Crude empirical scoring functions. [131]	Reconstruction of the ligand in the active site using rigid fragments, followed by minimization. [132]	Automatic placement of discrete interfacial waters. [133]	Manual or from the conformation of a bound reference ligand.	GUI or scripts.	Yes: PVM.	Commercial. Lease-type yearly license.	Very fast.
ICM-Dock [134-136]	Full.	Use of a rigid receptor conformations set, as input or generated [137] (MC simulation of complexes with known binders).	FF-based with additional components taking desolvation [138] and entropy effects into account.	Grid search in internal coordinates with minimization and MC optimization. [139]	Implicit (scoring function).	Can be defined manually; if not known, comprehensive graphical cavity detection wizards are provided.	Yes (ICM-Pro).	No.	Integrated into the ICM-Pro and ICM-VLS commercial packages.	The assessment of receptor flexibility seems a good trade-off between speed and accuracy.
Glide [140-142]	Search space restricted after first docking step.	No: receptor pre-processed as a grid. Cross-docking with multiple receptors possible using scripts.	Varying combinations of terms (depending on the docking process level) taken from an empiric scoring function and a FF.	Hierarchical: (1) selection of initial poses, (2) minimization, (3) MC refinement of the best conformers.	Explicit: bridging waters docked on the target grid along with the ligands.	Grid generated from center point of from the conformation of a known bound ligand.	Integrated in the Maestro GUI. [143]	Distributed calculation on clusters possible with special version.	Commercial. Maestro is free to academics.	Multiple-step procedure for improving the speed/precision ratio.
SHEF [144]	No.	No.	RMSD between two spherical harmonics surfaces (pure geometrical matching). [145, 146]	Minimization of the shape similarity function. [145]	No.	Receptor shape defined as set of spherical harmonics coefficients (same for all ligands), pre-processed by MSSH. [147]	Implemented in VSM-G [67], coupled with MSSH.	Planned in future VSM-G versions; will use APST [148-150].	Contact the authors.	Extremely fast, ideal as the first module in a multi-step VS protocol.
DARWIN [151]	Full.	No.	DARWIN is coupled with the popular CHARMM program [152].	GA-driven minimization.	Optional, implicit. Electrostatics treatment reported as hazardous.	Can be left to the program.	Implemented as an extension of CHARMM syntax.	Yes: PVM. Unix only.	Source code available on demand.	Readily usable as a VS procedure spawning CHARMM jobs.
MORDOR [153]	Full.	Full.	FF interaction energy.	Constrained MD associated with a minimizer.	?	Automatically detected as ligand forced to move along the protein surface.	?	Yes.	?	Full flexibility for both the protein and the ligand.

Table I: properties of the main used docking programs

References

1. Drews J. Drug discovery: A historical perspective. *Science* **287** (2000) 1960-1964.

2. Charifson P.S. *Practical application of computer-aided drug design*. (1997) Marcel Dekker, 552 pages.

3. Jorgensen W.L. The many roles of computation in drug discovery. *Science* **303**, issue 5665 (2004) 1813-1818.

4. James L.C. and Tawfik D.S. Conformational diversity and protein evolution - a 60-year-old hypothesis revisited. *Trends in Biochemical Sciences* **28**, issue 7 (2003) 361-368.

5. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. and Bourne P.E. The Protein Data Bank. *Nucleic Acids Research* **28**, issue 1 (2000) 235-242.

6. Smith Schmidt T. Banking on structures. *BioIT World* **1**, issue 8 (2002).

7. Chanda S.K. and Caldwell J.S. Fulfilling the promise: drug discovery in the post-genomic era. *Drug Discovery Today* **4** (2003) 168-174.

8. Nordhaus, W.D. *An economic history of computing* (2006).
http://nordhaus.econ.yale.edu/computing_June2006.pdf

9. Neshich G., Borro L.C., Higa R.H., Kuser P.R., Yamagishi M.E.B., Franco E.H., Krauchenco J.N., Fileto R., Ribeiro A.A., Bezerra G.B.P., Velludo T.M., Jimenez T.S., Furukawa N., Teshima H., Kitajima K., Bava A., Sarai A., Togawa R.C. and Mancini A.L. The Diamond STING server. *Nucleic Acids Research* **33**, issue Web server issue (2005) W29-W35.

10. Neshich G., Mancini A.L., Yamagishi M.E., Kuser P.R., Fileto R., Pinto I.P., Palandrani J.F., Krauchenco J.N., Baudet C., Montagner A.J. and Higa R.H. STING Report: convenient web-based application for graphic and tabular presentations of protein sequence, structure and function descriptors from the STING database. *Nucleic Acids Research* **33**, issue Database issue (2005) D269-D274.

11. STING - Sequence To and withIN Graphics.
http://www.cbi.cnptia.embrapa.br/SMS/

12. DiMasi J.A., Hansen R.W. and Grabowski H.G. The price of innovation: new estimates of drug development costs. *Journal of Health Economics* **22** (2003) 151-185.

13. Mervis J. Productivity counts - but the definition is key. *Science* **309** (2005) 726-727.

14. Vogel H.G., Rieß G. and Vogel W.F. Strategies in drug discovery and evaluation, in *Drug discovery and evaluation - pharmacological assays*, ed. H.G. Vogel (2002) Springer-Verlag, Berlin Heidelberg New-York, p. 1-21.

15. Anderson D. Public computing: Reconnecting people to science, in *Conference on shared knowledge and the Web* (2003) Madrid, Spain.

16. Rauwerda H., Roos M., Hertzberger B.O. and Breit T.M. The promise of a virtual lab in drug discovery. *Drug Discovery Today* **11**, issue 5-6 (2006) 228-236.

17. Garrett M.D. and Workman P. Discovering novel chemotherapeutic drugs for the third millenium. *European Journal of Cancer* **35**, issue 14 (1999) 2010-2030.

18. Searls D.B. Using bioinformatics in gene and drug discovery. *Drug Discovery Today* **5**, issue 4 (2000) 135-143.

19. Shneiderman B. Inventing discovery tools: combining information visualization with data mining. *Information Visualization* **1**, issue 1 (2002) 5-12.

20. Leach A.R. *Molecular modelling: Principles and applications*. (1996) Addison Wesley Longman, Essex, 585 pages.

21. Ooms F. Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Current Medicinal Chemistry* **7** (2000) 141-158.

22. Oprea T.I. and Matter H. Integrating virtual screening in lead discovery. *Current Opinion in Chemical Biology* **8**, issue 4 (2004) 349-358.

23. Lyne P.D. Structure-based virtual screening: an overview. *Drug Discovery Today* **7**, issue 20 (2002) 1047-1055.

24. Anderson A.C. The process of structure-based drug design. *Chemistry & Biology* **10** (2003) 787-797.

25. Veselovsky A.V. and Ivanov A.S. Strategy of computer-aided drug design. *Current Drug Targets - Infectious Disorders* **3**, issue 1 (2003) 33-40.

26. Kuntz I.D., Blaney J.M., Oatley S.J., Langridge R. and Ferrin T.E. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology* **161**, issue 2 (1982) 269-288.

27. Goodford P.J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of Medicinal Chemistry* **28** (1985) 849-857.

28. Gschwend D.A., Good A.C. and Kuntz I.D. Molecular docking towards drug discovery. *Journal of Molecular Recognition* **9** (1996) 175-186.

29. Brooijmans N. and Kuntz I.D. Molecular recognition and docking algorithms. *Annual Review of Biophysics and Biomolecular Structure* **32** (2003) 335-373.

30. Kitchen D.B., Decornez H., Furr J.R. and Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery* **3** (2004) 935-949.

31. Hilbert M., Böhm G. and Jaenicke R. Structural relationships of homologous proteins as a fundamental principle in homology modeling. *Proteins: Structure, Function, and Genetics* **17**, issue 2 (1993) 138-151.

32. Kubinyi H. QSAR and 3D QSAR in drug design part 1: Methodology. *Drug Discovery Today* **2**, issue 11 (1997) 457-467.

33. Kubinyi H. QSAR and 3D QSAR in drug design part 2: Applications and problems. *Drug Discovery Today* **2**, issue 12 (1997) 538-546.

34. Forrest S. Genetic algorithms: principles of natural selection applied to computation. *Science* **261** (1993) 872-878.

35. Schneider G. and Wrede P. Artificial neural networks for computer-based molecular design. *Progress in Biophysics and Molecular Biology* **70**, issue 3 (1998) 175-222.

36. Zapan J. and Gasteiger J. *Neural networks in chemistry and drug design*. 2nd ed. (1999) Wiley-VCH, 402 pages.

37. Walters W.P., Stahl M.T. and Murcko M.A. Virtual screening - an overview. *Drug Discovery Today* **3**, issue 4 (1998) 160-178.

38. Bissantz C. *Development and application of new methods for the virtual screening of chemical databases.* Thesis (2002) Swiss Federal Institute of Technology, Zurich, 292 pages.

39. Shoichet B.K. Virtual screening of chemical libraries. *Nature* **432** (2004) 862-865.

40. Mestres J. Virtual screening: a real screening complement to high-throughput screening. *Biochemical Society Transactions* **30**, issue 4 (2002) 797-799.

41. von Itzstein M., Wu W.-Y., Kok G.B., Pegg M.S., Dyason J.C., Jin B., van Phan T., Smythe M.L., White H.F., Oliver S.W., Colman P.M., Varghese J.N., Ryan D.M., Woods J.M., Bethell R.C., Hotham V.J., Cameron J.M. and Penn C.R. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363** (1993) 418-423.

42. Schevitz R.W., Bach N.J., Carlson D.G., Chirgadze N.Y., Clawson D.K., Dillard R.D., Draheim S.E., Hartley L.W., Jones N.D., Mihelich E.D., Olkowski J.L., Snyder D.W., Sommers C. and Wery J.-P. Structure-based design of the first potent and selective inhibitor of human non-pancreatic secretory phospholipase A2. *Nature Structural Biology* **2** (1995) 458-465.

43. Tondi D., Slomczynska U., Costi M.P., Watterson D.M., Ghelli S. and Shoichet B.K. Structure-based discovery and in-parallel optimization of novel competitive inhibitors of thymidylate synthase. *Chemistry & Biology* **6**, issue 5 (1999) 319-331.

44. Filikov A.V., Mohan V., Vickers T.A., Griffey R.H., P.D C., Abagyan R.A. and James T.L. Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *Journal of Computer-Aided Molecular Design* **14**, issue 6 (2000) 593-610.

45. Hopkins S.C., Vale R.D. and Kuntz I.D. Inhibitors of kinesin activity from structure-based computer screening. *Biochemistry* **39**, issue 10 (2000) 2805-2814.

46. Perola E., Xu K., Kollmeyer T.M., Kaufmann S.H., Prendergast F.G. and Pang Y.P. Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *Journal of Medicinal Chemistry* **43**, issue 3 (2000) 401-408.

47. Doman T.N., McGovern S.L., Witherbee B.J., Kasten T.P., Kurumbail R., Stallings W.C., Conolly D.T. and Shoichet B.K. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *Journal of Medicinal Chemistry* **45** (2002) 2213-2221.

48. Grüneberg S., Stubbs M.T. and Klebe G. Successful virtual screening for novel inhibitors of human carbonic anhydrase: Strategy and experimental confirmation. *Journal of Medicinal Chemistry* **45** (2002) 3588-3602.

49. Shoichet B.K., McGovern S.L., Wei B. and Irwin J.J. Lead discovery using molecular docking. *Current Opinion in Chemical Biology* **6** (2002) 439-446.

50. Vangrevelinghe E., Zimmermann K., Schoepfer J., Portmann R., Fabbro D. and Furet P. Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. *Journal of Medicinal Chemistry* **46**, issue 13 (2003) 2656-2662.

51. Kraemer O., Hazemann I., Podjarny A.D. and Klebe G. Virtual screening for inhibitors of human aldose reductase. *Proteins: Structure, Function, and Bioinformatics* **55** (2004) 814-823.

52. Lahana R. How many leads from HTS? *Drug Discovery Today* **4**, issue 10 (1999) 447-448.

53. Golebiowski A., Klopfenstein S.R. and Portlock D.E. Lead compounds discovered from libraries. *Current Opinion in Chemical Biology* **5**, issue 3 (2001) 273-284.

54. Golebiowski A., Klopfenstein S.R. and Portlock D.E. Lead compounds discovered from libraries: part 2. *Current Opinion in Chemical Biology* **7**, issue 3 (2003) 308-325.

55. Shoichet B.K. Screening in a spirit haunted world. *Drug Discovery Today* **11**, issue 13-14 (2006) 607-615.

56. Beddell C.R., Goodford P.J., Norrington F.E., Wilkinson S. and Wootton R. Compounds designed to fit a site of known structure in human haemoglobin. *British Journal of Pharmacology* **57**, issue 2 (1976) 201-209.

57. PDB - RCSB Protein Data Bank. http://www.rcsb.org/pdb/home/home.do

58. Kellenberger E., Rodrigo J., Muller P. and Rognan D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Structure, Function, and Bioinformatics* **57** (2004) 225-242.

59. Kontoyianni M., McClellan L.M. and Sokol G.S. Evaluation of docking performance: comparative data on docking algorithms. *Journal of Medicinal Chemistry* **47**, issue 3 (2004) 558-565.

60. Perola E., Walters W.P. and Charifson P.S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins: Structure, Function, and Bioinformatics* **56**, issue 2 (2004) 235-249.

61. Kubinyi H. Drug research: myths, hype and reality. *Nature Reviews Drug Discovery* **2** (2003) 665-668.

62. Verkhivker G.M., Bouzida D., Gehlhaar D.K., Rejto P.A., Arthurs S., Colson A.B., Freer S.T., Larson V., Luty B.A., Marrone T. and Rose P.W. Deciphering common failures in molecular docking of ligand-protein complexes. *Journal of Computer-Aided Molecular Design* **14** (2000) 731-751.

63. Clarke C., Woods R.J., Gluska J., Cooper A., Nutley M.A. and Boons G.-J. Involvement of water in carbohydrate-protein binding. *Journal of the American Chemical Society* **123** (2001) 12238-12247.

64. Ladbury J.E. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chemistry & Biology* **3**, issue 12 (1996) 973-980.

65. Charifson P.S., Corkery J.J., Murcko M.A. and Walters W.P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *Journal of Medicinal Chemistry* **42**, issue 25 (1999) 5100-5109.

66. Feher M. Consensus scoring for protein-ligand interactions. *Drug Discovery Today* **11**, issue 9-10 (2006) 421-428.

67. Beautrait A., Leroux V., Chavent M., Maigret B., Cai W., Shao W., Moreau G., Bladon P., Yao J., Liao Q., Yu F. and Souchet M. VSM-G: the Virtual Screening Manager platform for computational Grids. Example of use for the identification of putative liver X receptor ligands. *Manuscript in preparation* (2006).

68. Dobson C.M. Chemical space and biology. *Nature* **432** (2004) 824-828.

69. Newton C.G. and Lockey P.M. The importance of early pharmacokinetics. *Current Drug Discovery* **April 2003** (2003) 33-36.

70. Sadowski J. Optimization of the drug-likeness of chemical libraries. *Perspectives in Drug Discovery and Design* **20** (2000) 17-28.

71. Irwin J.J. How good is your screening library? *Current Opinion in Chemical Biology* **10**, issue 4 (2006) 352-356.

72. Lipinski C.A., Lombardo F., Dominy B.W. and Feeney P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Discovery Reviews* **23**, issue 1 (1997) 3-25.

73. Irwin J.J. and Shoichet B.K. ZINC - a free database of comercially available compounds for virtual screening. *Journal of Chemical Information and Modeling* **45**, issue 1 (2005) 177-182.

74. A free database for virtual screening: ZINC - Zinc Is Not Commercial. http://blaster.docking.org/zinc/

75. Bleicher K.H., Böhm H.-J., Müller K. and Alanine A.I. Hit and lead generation: beyond high-throughput screening. *Nature Reviews Drug Discovery* **2**, issue 5 (2003) 369-378.

76. Bajorath J. Integration of virtual and high-throughput screening. *Nature Reviews Drug Discovery* **1** (2002) 882-894.

77. Alder B.J. and Wainwright T.E. Studies in molecular dynamics. I. General method. *Journal of Chemical Physics* **31**, issue 2 (1959) 459-466.

78. van Gunsteren W.F. and Berendsen J.C. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angewandte Chemie - International edition in english* **29** (1990) 992-1023.

79. Karplus M. and McCammon J.A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **9**, issue 9 (2002) 646-652.

80. Chipot C. and Pearlman D.A. Free energy calculations: the long and winding gilded road. *Molecular Simulation* **28** (2002) 1-12.

81. Pearlman D.A., Case D.A., Caldwell J.W., Ross W.S., Cheatham III T.E., DeBolt S., Ferguson D., Seibel G. and Kollman P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **91** (1995) 1-41.

82. Nelson M., Humphrey W., Gursoy A., Dalke A., Kalé L., Skeel R. and Schulten K. NAMD - A parallel, object-oriented molecular dynamics program. *International Journal of Supercomputer Applications and High Performance Computing* **10** (1996) 251-268.

83. MacKerell Jr. A.D., Brooks B., Brooks III C.L., Nilsson L., Roux B., Won Y. and Karplus M. CHARMM: The energy function and its parametrization with an overview of the program, in *The Encyclopedia of computational chemistry*, ed. P.V.R. Schleyer, et al. (1998) John Wiley & sons, Chichester, p. 271-277.

84. Kalé L., Skeel R., Bhandarkar M., Brunner R., Gursoy A., Krawetz N., Phillips J., Shinozaki A., Varadarajan K. and Schulten K. NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics* **151** (1999) 283-312.

85. Case D.A., Cheatham III T.E., Darden T., Gohlke H., Luo R., Merz Jr. K.M., Onufriev A., Simmerling C., Wang B. and Woods R.J. The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **26**, issue 16 (2005) 1668-1688.

86. Johnson M.A., Galván I.F. and Villà-Freixa J. Framework-based design of a new all-purpose molecular simulation application: The Adun simulator. *Journal of Computational Chemistry* **26**, issue 15 (2005) 1647-1659.

87. Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kalé L. and Schulten K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, issue 16 (2005) 1781-1802.

88. van der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A.E. and Berendsen H.J.C. GROMACS: fast, flexible and free. *Journal of Computational Chemistry* **26**, issue 16 (2005) 1701-1718.

89. Adun molecular simulation project. http://diana.imim.es/Adun

90. Amber - Assisted Model Building with Energy Refinement. http://amber.scripps.edu/

91. CHARMm - Chemistry at HARvard Molecular Mechanics. http://www.accelrys.com/products/charmm/index.html

92. GROMACS - GROningen MAchine for Chemical Simulation. http://www.gromacs.org/

93. NAMD - Not Another Molecular Dynamics. http://www.ks.uiuc.edu/Research/namd/

94. Mason J.S., Good A.C. and Martin E.J. 3-D pharmacophores in drug discovery. *Current Pharmaceutical Design* **7**, issue 7 (2001) 567-597.

95. Cho A.E., Guallar V., Berne B.J. and Freisner R. Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) approach. *Journal of Computational Chemistry* **26**, issue 9 (2005) 915-931.

96. Freisner R.A., Murphy R.B., Repasky M.P., Frye L.L., Greenwood J.R., Halgren T.A., Sanschagrin P.C. and Mainz D.T. Extra precision Glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of Medicinal Chemistry* **49**, issue 21 (2006) 6177-6196.

97. Swiss-Prot Protein knowledgebase. http://www.expasy.org/sprot/

98. Allen F.H., Bellard S., Brice M.D., Cartwright B.A., Doubleday A., Higgs H., Hummelink T., Hummelink-Peters B.G., Kennard O., Motherwell W.D.S., Rodgers J.R. and Watson D.G. The Cambridge Crystallographic Data Centre: Computer-based search, retrieval, analysis and display of information. *Acta Crystallographica B* **35** (1979) 2331-2339.

99. Bairoch A. Serendipity in bioinformatics, the tribulations of a Swiss bioinformatician through exciting times! *Bioinformatics* **16**, issue 1 (2000) 48-64.

100. Allen F.H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallographica B* **58** (2002) 380-388.

101. Boeckmann B., Bairoch A., Apweiler R., Blatter M.-C., Estreicher A., Gasteiger E., Martin M.J., Michoud K., O'Donovan C., Phan I., Pilbout S. and Schneider M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research* **31**, issue 1 (2003) 365-370.

102. Yang G., Desvignes M.D., Smaïl-Tabone M. and Maigret B. TLdb: Target-Ligand database. **Manuscript in preparation** (2006).

103. CSD - Cambridge Structural Database. http://www.ccdc.cam.ac.uk/products/csd/

104. Stoica I., Morris R., Karger D., Kaashoek M.F. and Balakrishnan H. Chord: a scalable peer-to-peer lookup service for internet applications, in *Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications* (2001).

105. Anderson D.P., Cobb J., Korpela E., Lebofsky M. and Werthimer D. SETI@home: An experiment in public-resource computing. *Communications of the ACM* **45**, issue 11 (2002) 56-61.

106. Chen S., Zhang W., Ma F. and Shen J. A cooperative computing platform for drug discovery and design, in *Proceedings of the 2004 IEEE international conference on services computing* (2004).

107. Montgomery S.B., Fu T., Guan J., Lin K. and Jones S.J. An application of peer-to-peer technology to the discovery, use and assessment of bioinformatics programs. *Nature Methods* **2**, issue 8 (2005) 563.

108. Cappello F., Caron E., Dayde M., Desprez F., Jegou Y., Primet P., Jeannot E., Lanteri S., Leduc J., Melab N., Mornet G., Namyst R., Quetier B. and Richard O. Grid'5000: a large scale and highly reconfigurable grid experimental testbed, in *The 6th IEEE/ACM International Workshop on Grid Computing* (2005).

109. Coveney P.V. Scientific grid computing. *Philosophical Transactions of the Royal Society of London Series A - Mathematical, Physical and Engineering Sciences* **363**, issue 1833 (2005) 1707-1713.

110. Woods C.J., Ng M.H., Johnston S., Murdock S.E., Wu B., Tai K., Fangohr H., Jeffreys P., Cox S., Frey J.G., Sansom M.S.P. and Essex J.W. Grid computing and biomolecular simulation. *Philosophical Transactions of the Royal Society of London Series A - Mathematical, Physical and Engineering Sciences* **363**, issue 1833 (2005) 2017-2035.

111. Grid'5000. http://www.grid5000.fr

112. Paolini G.V., Shapland R.H.B., van Hoorn W.P., Mason J.S. and Hopkins A.L. Global mapping of pharmacological space. *Nature Biotechnology* **24**, issue 7 (2006) 805-815.

113. Moustakas D.T., Lang P.T., Pegg S., Pettersen E.T., Kuntz I.D., Brooijmans N. and Rizzo R.C. Development and validation of a modular, extensible docking program: DOCK 5. *Submitted for publication* (2006).

114. UCSF DOCK. http://dock.compbio.ucsf.edu/

115. Ewing T.J.A., Makino S., Skillman A.G. and Kuntz I.D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *Journal of Computer-Aided Molecular Design* **15** (2001) 411-428.

116. Gropp W., Lusk E. and Skjellum A. *Using MPI: Portable parallel programming with the message-passing interface*. (1994) MIT Press, 328 pages.

117. Gropp W., Lusk E., Doss N. and Skjellum A. A high-performance, portable implementation of the MPI message passing interface standard. *Parallel Computing* **22**, issue 6 (1996) 789-828.

118. Jones G., Willett P. and Glen R.C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology* **245**, issue 1 (1995) 43-43.

119. Jones G., Willett P., Glen R.C., Leach A.R. and Taylor R. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* **267** (1997) 727-748.

120. GOLD - Genetic Optimization for Ligand Docking. http://www.ccdc.cam.ac.uk/products/life_sciences/gold/

121. Verdonk M.L., Chessari G., Cole J.C., Hartshorn M.J., Murray C.W., Nissink J.W.M., Taylor R.D. and Taylor R. Modeling water molecules in protein-ligand docking using GOLD. *Journal of Medicinal Chemistry* **48** (2005) 6504-6515.

122. Geist A., Beguelin A., Dongarra J., Jiang W., Manchek R. and Sunderam V.S. *PVM: Parallel Virtual Machine - A users' guide and tutorial for network parallel computing*. (1994) MIT Press, 299 pages.

123. Sunderam V.S. PVM: A framework for parallel distributed computing. *Concurrency, Practice and Experience* **2**, issue 4 (1990) 315-340.

124. Goodsell D.S., Morris G.M. and Olson A.J. Automated docking of flexible ligands: applications of AutoDock. *Journal of Molecular Recognition* **9**, issue 1 (1996) 1-5.

125. AutoDock. http://autodock.scripps.edu/

126. Österberg F., Morris G.M., Sanner M.F., Olson A.J. and Goodsell D.S. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins: Structure, Function, and Genetics* **46**, issue 1 (2001) 34-40.

127. Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., Belew R.K. and Olson A.J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* **19**, issue 14 (1998) 1639-1662.

128. Rarey M., Kramer B., Lengauer T. and Klebe G. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology* **261**, issue 3 (1996) 470-489.

129. FlexX. http://www.biosolveit.de/FlexX/

130. Claußen H., Buning C., Rarey M. and Lengauer T. FlexE: efficient molecular docking considering protein structure variations. *Journal of Molecular Biology* **308**, issue 2 (2001) 377-395.

131. Stahl M. and Rarey M. Detailed analysis of scoring functions for virtual screening. *Journal of Medicinal Chemistry* **44**, issue 7 (2001) 1035-1042.

132. Hoffmann D., Kramer B., Washio T., Steinmetzer T., Rarey M. and Lengauer T. Two-stage method for protein-ligand docking. *Journal of Medicinal Chemistry* **42**, issue 21 (1999) 4422-4433.

133. Rarey M., Kramer B. and Lengauer T. The particle concept: Placing discrete water molecules during protein-ligand docking predictions. *Proteins: Structure, Function, and Bioinformatics* **34** (1999) 17-28.

134. Cavasotto C.N. and Abagyan R.A. Protein flexibility in ligand docking and virtual screening to protein kinases. *Journal of Molecular Biology* **337** (2004) 209-225.

135. Bursulaya B.D., Totrov M., Abagyan R. and Brooks III C.L. Comparative study of several algorithms for flexible ligand docking. *Journal of Computer-Aided Molecular Design* **17** (2003) 755-763.

136. MolSoft. http://www.molsoft.com

137. Cavasotto C.N., Kovacs J.A. and Abagyan R.A. Representing receptor flexibility in ligand docking through relevant normal modes. *Journal of the American Chemical Society* **127**, issue 26 (2005) 9632-9640.

138. Fernandez-Recio J., Abagyan R. and Totrov M. Improving CAPRI predictions: optimized desolvation for rigid-body docking. *Proteins: Structure, Function, and Bioinformatics* **60**, issue 2 (2005) 308-313.

139. Totrov M. and Abagyan R. Flexible protein-ligand docking by docking energy optimization in internal coordinates. *Proteins: Structure, Function, and Genetics* **Suppl. 1** (1997) 215-220.

140. Freisner R.A., Banks J.L., Murphy R.B., Halgren T.A., Klicic J.J., Mainz D.T., Repasky M.P., Kmoll E.H., Shelley M., Perry J.K., Shaw D.E., Francis P. and Shenkin P.S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry* **47**, issue 7 (2004) 1739-1749.

141. Halgren T.A., Murphy R.B., Freisner R.A., Beard H.S., Frye L.L., Pollard W.T. and Banks J.L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of Medicinal Chemistry* **47**, issue 7 (2004) 1750-1759.

142. Glide - Grid-based LIgand Docking with Energetics. http://www.schrodinger.com/ProductDescription.php?mID=6&sID=6

143. Maestro - Unified interface for Schrödinger software. http://www.schrodinger.com/ProductDescription.php?mID=6&sID=15

144. Cai W., Xu J., Shao X. and Maigret B. SHEF: An efficient approach for virtual screening using coefficients of spherical harmonics surfaces. *Submitted for publication* (2006).

145. Ritchie D.W. and Kemp G.J.L. Fast computation, rotation and comparison of low resolution spherical harmonic molecular surfaces. *Journal of Computational Chemistry* **20**, issue 4 (1999) 383-395.

146. Cai W., Shao X. and Maigret B. Protein-ligand recognition using spherical harmonic molecular surfaces: towards a fast and efficient filter for large virtual throughput screening. *Journal of Molecular Graphics and Modelling* **20**, issue 4 (2002) 313-328.

147. Yamagishi M.E., Martins N.F., Neshich G., Cai W., Shao X., Beautrait A. and Maigret B. A fast surface-matching procedure for protein-ligand docking. *Journal of Molecular Modeling* **12**, issue 6 (2006) 965-972.

148. Casanova H. and Berman F. Parameter sweeps on the grid with APST, in *Grid Computing*, eds. F. Berman, G. Fox, and T. Hey (2003) John Wiley & Sons, p. 773-787.

149. van der Raadt K., Yang Y. and Casanova H. Practical divisible load scheduling on grid platforms with APST-DV, in *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05) - Papers* (2005).

150. APST - A Parameter Sweep Tool. http://grail.sdsc.edu/projects/apst/

151. Taylor J.S. and Burnett R.M. DARWIN: A program for docking flexible molecules. *Proteins: Structure, Function, and Genetics* **41**, issue 2 (2000) 173-191.

152. Brooks B.R., Bruccoleri R.E., Olafson B.D., States D.J., Swaminathan S. and Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4**, issue 2 (1983) 187-217.

153. MORDOR - MOlecular Recognition with a Driven dynamics OptimizeR. http://mondale.ucsf.edu/science/mordor.html



《计算机与应用化学》征稿简则

《计算机与应用化学》为中国科学院主管学术性刊物，系国家级化学化工类中文核心期刊，被评为中国科学院优秀期刊、全国石油和化工行业优秀期刊、中国期刊方阵期刊，并被国内外重要检索刊物和机构收录。国内外公开发行。

本刊重点刊登将先进的计算方法和信息技术运用于化学及相关过程的原创性研究论文以及评述性文章，传播相关知识领域的技术和方法，主要内容包括但不限于分子模型化、过程模拟与系统集成、信息系统、化学计量学、计算机辅助教学等。

本刊栏目专题为计算机化学、分子模拟与分子设计、药物分子设计、材料科学、生物信息与技术、信息技术与化学化工应用、石油与精细化工、安全技术、化学相关过程、化学信息学、应用与绿色化学、分析化学、计算机辅助教学等。栏目类型为综述、研究论文、研究快报、学术论坛、应用技术、软件介绍、信息交流、书讯、讲座、广告(人才需求、专业类需求与推广、产品介绍)。

● 本刊将快速报道前沿性学科高水平的成果，及时刊登新理论、新工艺、高新技术的原始创新文章。

● 本刊鼓励发表英文稿件，凡学术水平、文字水平较高或由多国作者合作，以及国外作者发表的文章，均可用英文发表。欢迎国内外及港澳台地区学者投稿。

来稿要求： 作者须遵守《中华人民共和国著作权法》，不得一稿多投，引用他人作品时，务必在参考文献中著录，严禁抄袭，文责自负。来稿应论点明确，文字精炼，数据可靠，符号正确。

属于研究生的稿件要求由导师审查后在稿件上签署姓名和日期。

属于课题研究项目和涉及密级内容的文章，需由作者单位出据证明。

注意事项： 稿件要求标题具有描述性，摘要以第三人称写明文章特点，以及研究目的、具体方法、结果和结论；作者 1-5 人；作者单位全称，地名，邮政编码；中国图书分类法号码(在图书馆查阅)；必须有中英文篇名、作者、单位全称、图表题目，国家、省、部、市级和国外资助项目名称及编号，参考文献(作者、文章名、(书)刊名、年卷期起止页码)，参考文献必须是英文，如果是中文参考文献要求中英文对照，引用的参考文献必须查原始文献；重点论文图表中的文字均为英文；稿件页数以双页为宜。

投稿时需注明详细联系地址、电话及 Email。

附注事项： 投稿请平信寄 1 份打印稿，同时交纳 100 元注册费，稿件录用后按文章内容交纳发表费，缴送 .DOC 文件和原始 .TIF 或 .JPG 文件，并寄 1 份打印稿(稿件行距略宽)。

广告办理事项： 请携带法人营业执照(副本)及复印件(2 份)、法人身份证及复印件(2 份)，行业许可证及复印件。个人求职求学广告业务请携带身份证及复印件。

广告收费标准： 每期制作费：黑白整版 800 元，彩色整版 2000—3000 元；印刷费：黑白整版 3000 元，彩色整版 8000 元。条目类全价 200 元，相同内容多期刊登需增加材料费用。

来稿请寄《计算机与应用化学》编辑部，王红(收)，邮政编码：100080

地址：北京市中关村北二街 2 条 1 号(中国科学院过程工程研究所 38 号信箱)。

每周一上午 10:00-12:00 值班电话：086-010-62558482，Email：jshx@home.ipe.ac.cn

热线电话：086-010-60143754(小灵通)

本期编委：邵学广，潘忠孝，陆文聪，徐峻

协办单位：中国科学技术大学、上海大学、

中国化学会计算机化学专业委员会、中国化工学会信息技术应用专业委员会

计 算 机 与 应 用 化 学
(月刊) (1984 年创刊)
2007 年 第 24 卷 第 1 期

Computers and Applied Chemistry
(Monthly) (Started in 1984)
Vol. 24 No. 1 2007

主管：中国科学院
主办：中国科学院过程工程研究所
主编：温浩
执行主编：王红
编辑：张涛
出版发行：《计算机与应用化学》编辑部
地址：北京中关村北二街 2 条 1 号，邮政编码：100080
电话：0086-010-62558482 E-mail：jshx@home.ipe.ac.cn
印刷：北京科文天和印刷有限公司
国内发行：全国各地邮电局
国外发行：中国国际图书贸易总公司中文报刊科
广告经营许可证号：京海工商广字第 0141 号

Chief Editor: Wen Hao
Executive Editor: Wang Hong
Editor: Zhang Tao
Published by Editorial Board of Computers and Applied Chemistry
Address: 1 Zhongguancun Beiertiao, Haidian District, Beijing 100080, China
Tel. 0086-010-62558482 E-mail: jshx@home.ipe.ac.cn
Printed by Beijing Kewen-Tianhe Printing Corporation Ltd.
Domestic Distributed by Local Post Offices
Abroad Distributed by China International Book Trading Corporation; Order by The Editorial

国内外公开发行 刊号：ISSN 1001-4160 / CN 11-3763/TP 国外发行代号：6346BM 国内邮发代号：82-500 每期征订价：33.00 元 零售价：35.00 元